Scale and Performance in Active Management are Not Negatively Related♣

John Adams, University of Texas at Arlington, jcadams@uta.edu

Darren Hayunga, University of Georgia, hayunga@uga.edu

Sattar Mansi, Virginia Tech, smansi@vt.edu

ABSTRACT

We revisit the nature of returns to scale following Pástor, Stambaugh, and Taylor (2015). Using replicated versions of their domestic equity fund sample, we confirm their negative and significant relation between industry scale and performance. However, upon closer examination we find the diseconomies of scale at the industry level result is an artifact of data errors that comprise less than 0.05% of the sample—168 out of 332,516 observations—that occurred most often in the year 2000. We are unable to find industry level diseconomies of scale in the post 2001 era. A major source of these errors is the incorrect use of Morningstar's current performance benchmarks to measure historical return performance. We confirm the non-result findings using Fama-French three-factor adjusted returns, which are not subject to benchmarking errors.

Keywords: Returns to scale, Active and passive management, Data errors, Index funds, Influential observations, Replication

JEL Specifications: G10, G11, G12

August 16, 2021

♣ We thank David Diltz, Nataliya Gerasimova, Qing Hao, Lubos Pástor, David Rakowski, Stephanie Rasmussen, Salil Sarkar, Terry Skantz, Robe Stambaugh, Thomas Thompson, Ram Venkataraman, Raghu Venugopalan, Sriram Villupuram, Ivo Welch (the editor), an anonymous referee, and seminar participants at the University of Texas Arlington for comments and suggestions that help improved the paper. Special thanks to Lucien Taylor and Min Zhu for providing software code as well as Arati Kale and Jason Morrison for data assistance. The remaining errors are ours.

There is a long-running debate over whether diseconomies of scale exist in the actively managed mutual fund industry.[1] This debate centers on the idea that money flows disproportionately into actively managed funds with the highest past returns. In their seminal work, 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 (e.g., additional resources to identify mispriced securities, bid-ask spreads, and price impacts from large trades). According to their model, there is an optimal level of actively managed assets and the excess is invested in a passive benchmark. As the fund grows, the portion that is passively managed increases and a fund eventually earns an expected return equal to that of its passive benchmark. In a similar vein, Pástor and Stambaugh (2012) argue that as money flows into actively managed funds, competition among funds for investment opportunities increases so that security prices are affected and it becomes more difficult for fund managers to beat passive indexes.

Early empirical research finds evidence of diseconomies of scale at the fund level. Chen et al. (2004) and Yan (2008), for instance, find that fund size has an adverse effect on return performance that is stronger among small-cap or less liquid funds. They argue that increased inflows to funds with illiquid holdings increase trading costs and price pressure on stocks, which subsequently lowers fund returns. One of the main criticisms of this research is the endogeneity of fund size. High-skilled managers tend to run larger funds, which by itself produces a positive performance-size relation across funds. This positive relation can completely offset any negative relation originating from diseconomies of scale. Put differently, skill is an omitted variable that is correlated with fund size biasing the performance-scale slope upwards. Recognizing this identification challenge, Pástor, Stambaugh, and Taylor (2015) (hereafter: PST) develop a recursive demeaning procedure that corrects for the omitted variable problem and find evidence of decreasing returns to scale but only at the industry level. More recently, Zhu (2018) modifies their recursive-demeaning estimator and finds evidence supporting fund-level diseconomies of scale.[2]

We examine the nature of returns to scale starting with PST. We focus on their work because of its contribution to the literature especially in addressing the endogeneity of manager skill, the availability of their code, and their meticulous sample documentation. Our motivation stems from concerns about methodological issues, variable construction, and sampling errors. PST utilize the

[1] Evidence supporting diseconomies of scale at the fund level include Yan (2008), Busse et al. (2021), Golez and Shive (2015), Harvey and Liu (2017), and Zhu (2018). Evidence of constant returns to fund scale include Edelen et al. (2007), Elton et al. (2012), Ferreira et al. (2013), Reuter and Zitzewitz (2015), and Phillips et al. (2018).

[2] Zhu (2018) result supporting industry level diseconomies of scale was not reported (see Appendix A.2 in her paper).

merged CRSP-Morningstar dataset to improve data accuracy and provide performance benchmarks. However, this approach only considers errors in the raw fund returns and size variables. Because errors are not limited to returns and size data, we use a multivariate outlier detection strategy to help identify potential data errors. It is well known that the OLS estimator, with and without fixed effects, is sensitive to extreme observations. While the recursive demeaning instrumental variable estimator remedies the finite sample bias in fixed effects models, it can exacerbate the influence of multivariate outliers including those arising from data errors.[3] As a single OLS regression is sensitive to outliers, it follows that in the instrumental variable regression these observations can influence both the first- and second-stage least squares estimates. Additionally, the recursive demeaning estimator employs the mean as a location metric, which is also sensitive to outliers.

Using replicated versions of their domestic equity fund sample, we confirm the main PST results by documenting a negative and significant relation between industry scale and performance and an insignificant relation at the fund level. Upon closer examination, however, the findings of diseconomies of scale at the industry level stem from extreme influential observations that constitute less than 0.05%—168 out of 332,516 observations—of the overall sample. A manual examination of these influential observations with verifications against outside sources including fund webpages, Morningstar Direct, prospectuses, annual reports, SEC filings, and news articles shows these observations are data errors that occurred most often in the year 2000 during the dot-com stock market crash.

We find that these data errors originate from sampling errors in the form of non-domestic and non-equity funds, omitted variables, and measurement errors due to incorrect benchmarks used to obtain alphas. The majority of the literature examining economies of scale use Fama-French (1996) factors to measure returns in excess of risk. However, as Berk and van Binsbergen (2017) and PST note, the Fama-French factors are long-short portfolios whose returns cannot be costlessly earned by mutual fund managers. In addition, Cremers et al. (2013) find that Fama-French models yield significant nonzero alphas for passive index portfolios. PST follow their recommendation and use Morningstar index-based benchmarks.

Morningstar provides separate data points for current and historical style categories. However, the Morningstar data points that PST use appear to be for the most recent period and do not account for

[3] For the purpose of this research, the term multivariate outliers refer to data points that do not fit the overall pattern of the sample and thereby heavily influence regression coefficients. Extreme and influential observations are outliers whose removal changes the estimated coefficients and standard errors and are potentially bad data.

historical variations in fund style. Because fund styles change over time and current benchmarks are often different from historical benchmarks, the benchmark-adjusted returns for some funds in their sample are measured with large errors. The comparison to benchmark indices introduces additional problems. First, fund styles change over time and even historical benchmarks become stale as Morningstar only periodically updates assigned benchmarks. Second, benchmarks do not account for funds whose investment strategies are not constrained to a single asset class or investment category. Thus, the benchmark does not reflect the style or risk of the fund. Third, there are no unique index-based benchmarks for some funds. For example, in the middle of 2017 there were only 89 Morningstar benchmark indices to cover the 115 Morningstar categories. This means measurement errors in Morningstar Category benchmark-adjusted returns are possible even when the fund returns themselves are correct.

PST argue that OLS fixed effects regressions capture the relation between gross benchmark-adjusted returns and INDUSTRYSIZE better-than-recursive demeaning regressions. If so, 0.50% of the sample, or 1,672 observations, drive the full sample industry size diseconomies of scale result. We confirm that the PST OLS fixed effects results are due to erroneous observations by employing a multi-stage data cleaning procedure that uses historical Morningstar Category and Style assignments to account for benchmarking errors and manual investigations to remove sampling errors arising from incorrect Morningstar category and style classifications. We also confirm our findings using Fama-French three-factor adjusted returns. An advantage of factor-adjusted returns is that they are not subject to Morningstar benchmark measurement errors. Using the full sample, we find an insignificant relation between fund and industry levels of scale and Fama-French three factor adjusted return performance.

Additional analysis shows that the relation between benchmark-adjusted fund returns and industry size is fragile. When we remove between 0.05% and 0.50% of the fund-month observations displaying the greatest outlyingness, the relation between industry size and performance becomes insignificant. Moreover, with the removal of 1% of the influential observations, the results demonstrate increasing returns to scale. Thus, while bad data are driving the diseconomies of industry scale finding, for 99% of the data there is a positive relation between industry size and return performance.

We corroborate our findings following Zhu (2018), who argues the PST recursive demeaning procedure suffers from an inherent misspecification that is problematic in the fund size process. Zhu (2018) modifies their recursive-demeaning estimator by adding an intercept in the first-stage regression and finds evidence supporting fund-level diseconomies of scale. We use the modified estimator and follow her sample generation method fully with the exception that we begin with the cleaned CRSP-Morningstar merged database rather than the raw Morningstar database and use the PST sample start date. We find insignificant scale results at both the industry and fund levels when the dollar level of assets under management is the functional form of fund size. The results for log of the dollar level of assets under management are sensitive to outliers and inconsistent across size-based subsamples.

Our analysis shows that it possible to find some combination of variables, ways to measure size and return, sample period begin-and-end dates, sample selection criteria, and estimation technique that yields a negative scale result if data errors are ignored. This perhaps explains why the evidence in the literature on diseconomies of scale and return performance is so mixed. Rather than attempt to fit some subset of our results to existing theory or develop new theory, we propose an alternative explanation. The empirical methods in diseconomies of scale research are powerful, especially the PST and Zhu (2018) versions of recursive demeaning regressions, but when the data contain errors, these methods identify on chance rather than actual economic trends. After considering errors in the data and the fragility of the results, we fail to reject the null hypothesis of constant returns to fund level diseconomies of scale. At the industry level, the evidence strongly supports non-negative returns to scale and there is even some evidence of positive economies of scale.

Our findings of non-negative returns to scale may appear at odds with the Berk and Green (2004) rational expectations framework and the Pástor and Stambaugh (2012) view that the actively managed mutual fund industry cannot exist under non-negative returns to scale. Berk and Green's (2004) separation of fund total assets under management into active and passive components provides some insight (see page 1278 in their paper). Although trading commissions, bid-ask spreads, and price impacts from large trades are typically associated with the actively managed portion of fund assets, the passively managed portion also experiences these operating costs to some degree. Berk and Green (2004) assume these non-expense ratio costs are fixed. However, if these costs are decreasing in fund size for either the actively or passively managed component, the observed empirical relation between gross benchmark-adjusted returns and fund size can be negative, constant, or positive depending on which effect dominates: decreasing returns to scale for actively managed monies or increasing returns to scale for operating costs.

We directly test the hypothesis of positive non-expense-ratio economies of scale in passively managed funds using a sample of domestic equity index funds. After removing 621 out of 25,256 fund-months of the most influential observations, we find that scale at the fund and industry levels is positively related to gross benchmark-adjusted returns. A manual examination of each of these observations shows sample, benchmark, and omitted variable errors (e.g., foreign funds, asset allocation, enhanced index funds, misclassified funds). This finding provides support for increasing returns to scale for the passively managed component and helps shed light on the apparent puzzle. Even in the presence of diseconomies of scale in the actively managed portion of a portfolio, there are positive economies for the passively managed portion. The net result is constant or positive fund-level economies of scale that naturally extend to the industry level.

The net constant or positive returns to scale finding we observe in actively managed funds does not imply non-negative returns to scale in the actively managed component of actively managed funds. On the contrary, positive returns to scale in the passively managed component (i.e., operating efficiencies) coupled with constant or positive returns to scale in the combined active and passive fund components is consistent with diseconomies of scale in the actively managed component. In Berk and van Binsbergen's (2015) dynamic equilibrium setting, the amount of value added by managers captures their skill level and is the product of gross alpha and fund size. By using index funds to isolate the operational (i.e., passive) component of gross alpha, we demonstrate that manager ability to extract value from financial markets is not narrowly limited to active management skills. Our results suggest reductions in operating costs represent some portion of value added by skilled fund managers. It is also consistent with Adams et al.'s (2010) result of considerable variation in fund operating performance, and more recently with Crane and Crotty's (2018) finding that index fund managers appear to exhibit skill.

Our paper is related to the recent work of Adams et al. (2018), who use the Chen et al. (2004) study as a laboratory to identify and treat error-induced outliers. However, we differ from theirs in several important ways. First, we examine economies of scale in passive as well as active management. Second, Adams et al. (2018) do not consider non-expense ratio operating costs and the endogeneity in fund size so they tell us little about the nature of return to scale. In contrast, this paper accounts for the endogenous nature of size and manager skill. Third, we use the more recently available and higher quality cross-validated CRSP-Morningstar database. This combined dataset has been shown to mitigate errors in fund size and returns (e.g., PST).

We contribute to the literature on the nature of returns to scale in three important ways. First, we demonstrate the importance of high-quality data. Although recent advances in CRSP and Morningstar database quality are notable, numerous data errors remain. That data errors are common in these databases, especially in non-audited data components such as style and benchmark assignments, should come as no surprise as their primary intended users are investors and not researchers. Because data errors are not unique to these databases, our research has implications beyond the study of mutual funds and potential diseconomies of scale in active management. Large sample sizes are common in finance research, so it is difficult to manually examine all the data at the observation level. The influential observation detection method we employ identifies potential bad data points thereby greatly reducing the cost of manually verifying data quality.

Second, we show that mutual fund performance reflects both active management skill and operating costs, and that scale affects the actively and passively managed portfolio components differently. This finding is important because it changes how we conceptualize fund manager skill. Third, this research has practical implications for regulators, investors, and fund providers. For example, consider the marketing 12b-1 fees paid by investors to increase fund inflows and fund size. The Morningstar Direct database reports the average 12b-1 fee is around 33 basis points and industrywide totals almost $11 billion for 2018. Under the assumption of diseconomies of scale, 12b-1 fees represent an economically large agency issue, as they are deadweight costs borne by investors for the benefit of fund managers and distributors. However, our findings of net constant returns to scale for the combined active and passive components suggest the potential agency costs are small or nonexistent.

The remainder of the paper is organized as follows. Section 1 describes the sample and data methodology. Section 2 examines the role of influential observations in mutual fund data. Section 3 provides evidence on the nature of returns to scale using PST and Zhu (2018). Section 4 uses a rationale based on Berk and Green (2004) and examines the results using index funds only. Section 5 provides discussion and concluding remarks.

1 Sample Formation and Methodology

We construct a sample of actively managed U.S.-domiciled mutual funds from the historical CRSP and Morningstar databases following the PST Data Appendix available from Lubos Pástor's website. The sample starts in 1979 to coincide with their start date and ends in December 2014. Although the mutual fund databases are available through 2016, we end our sample in 2014 to allow two years for the error-correcting processes at the funds and data providers to play out. The CRSP and Morningstar historical databases are merged by matching funds on ticker, CUSIP, and name. Non-matched funds are dropped from the sample. This merging facilitates the comparison of reported returns and asset values across the two databases to improve data accuracy. Funds that are not long domestic equity are dropped using keyword searches of Morningstar category names and Morningstar primary prospectus benchmarks. These include bond, international, sector, long-short equity, commodities, and real estate funds. In total, funds from 73 Morningstar categories are removed from the sample.

1.1 Comparing Match Rates

Table 1 reports statistics of the merged CRSP-Morningstar sample. Similar to Berk and van Binsbergen (2015) and PST, we find observations where the monthly returns differ between CRSP and Morningstar databases. Panel A provides differences in the reported returns of the CRSP and Morningstar mutual fund databases for the matched funds. Column 1 lists return differences in ranges. Columns 2 and 3 show the number and percentage of matched funds in each return difference range as reported in the PST Data Appendix, and columns 4 and 5 report results for our merge using their sample period. Columns 6 and 7 report our replication results using the PST primary analysis sample period (1993–2011), and columns 8 and 9 report results for our main sample period (1993–2014). We select March 1993 as the start date for our primary sample because many funds only report assets under management quarterly or yearly before that date. This start data also coincides with the PST main sample.

Our 1979–2011 sample replication (columns 4 and 5) matches closely with the PST data. We do not expect exact matches because of ongoing data improvements in the CRSP and Morningstar databases. As in PST, the overwhelming majority of matched observations exhibit no return differences between the CRSP and Morningstar databases. In percentage terms our sample has a higher Do Not Differ rate than PST (91.1% versus 88.5%). The improved match rates in our sample extend to the other difference categories. For instance, 0.50% of the PST matches have return differences greater than 100 basis points while none of our matches have such large return differences. Because we follow the PST database reconciliation process, we attribute the diminished return differences to data-quality improvements in the CRSP and Morningstar databases.

The return differences are even smaller for our main sample (1993–2014), where we report 99% of the matched CRSP-Morningstar funds have return differences of less than 11 basis points. This rate is an improvement over the 96.9% rate in PST. They attempt to fix the return differences that are greater than 10 basis points by following Berk and van Binsbergen (2015). However, because of our sample's minimal return differences, we do not apply the Berk and van Binsbergen (2015) fixes as they could potentially introduce additional problems (e.g., errors in the fund dividends data in either database).

1.2 Comparing Sample Characteristics

For all samples we compute Morningstar benchmark-adjusted returns, fund size (FUNDSIZE), and industry size (INDUSTRYSIZE) following PST. Panel B of Table 1 presents descriptive statistics of the merged CRSP-Morningstar data for actively managed domestic equity funds (conversely, Panel A includes observations from all funds). Panel B reports mean, standard deviation, and percentile statistics for the PST overall sample (1979–2011), for our replication periods (1979–2011 and 1993–2011), and for our main sample period (1993–2014). We also report minimum and maximum values for our replication and main samples to show the full range in the return and scale measures.[4]

The benchmark-adjusted returns in PST and our replications are similar in terms of mean, standard deviation, and percentile values. However, our replication results indicate very large negative and positive returns when computing minimum and maximum values [PST do not report minimums and maximums]. Our main sample period has similar benchmark-adjusted returns and the same minimum and maximum values as the replication period samples. This indicates that extreme values occur more often during the PST main sample period (1993–2011) than during the three additional years of our sample (2012–2014).

FUNDSIZE is also similar in the published and replication samples (in both cases inflated to 2011 values). However, the replication sample statistics show the presence of very large funds as the maximum FUNDSIZE is about seven times larger than the 99^{th} percentile of FUNDSIZE. For our main sample, FUNDSIZE is inflated to 2014 values and, as with the other samples, shows a large dispersion in FUNDSIZE where the largest fund is more than 19,000 times larger than the smallest and where the mean-sized fund is much larger, by about 6 times, than the median-sized fund. Finally, Panel B reports considerable variations in INDUSTRYSIZE, which is consistent with PST.

1.3 Comparison with Published Regression Estimates

Table 2 provides replicated results using PST's main findings found in their Panel A of Table 3, where monthly portfolio-level gross Morningstar benchmark-adjusted fund returns are regressed on prior month FUNDSIZE and INDUSTRYSIZE. Table 2 compares the published results in columns 1, 2, and 3 with our replicated results in columns 4, 5, and 6.[5] The number of observations in our

---

[4] To match the regression results, we only include observations with non-missing values and with at least 24 months of data when computing the descriptive statistics.

[5] We apply the positive first-stage regression filter for forward and backward demeaned fund size as detailed in footnote 8 of PST. Hence, the discrepancy in the reported number of observations in Tables 1 and 2.

replications is slightly larger (by about 0.50%) than the number reported in PST. As before, we attribute this small increase to data improvements in the CRSP and Morningstar databases.

The estimated coefficients and *t*-statistics are similar. For the OLS specifications in columns 1 and 4, the estimated coefficients only differ at the third decimal place and the significance levels are comparable. In the OLS fixed effects specifications in columns 2 and 5, the published and our replicated results are also only minimally different. However, our main interest is in the recursive demeaning results as both the OLS and OLS fixed effects specifications suffer from omitted variable and finite-sample bias, respectively. In both the published and our replicated results, the estimated coefficient on fund size is insignificantly different from zero. Likewise, the estimated coefficients for INDUSTRYSIZE are quite close in magnitude and statistical significance. Table 2 shows the efficacy of our sample replication process and we move to the broader question of the effects, if any, of scale on performance.

2 Data Errors

2.1 Outlier Concerns

Influential data errors affect coefficient estimates, lead to Type I errors, and appear as outliers. Our concerns that data errors may be problematic stem from methodological issues, variable construction, and sampling errors. The OLS estimator, with or without fixed effects, is known to be sensitive to outliers. While the recursive demeaning instrumental variable estimator remedies the finite sample bias in fixed effects models, it amplifies the influence of outliers (e.g., Cohen-Freue et al., 2013). The intuition is straightforward. Because a single least squares regression is sensitive to outliers it follows that these observations can influence both the first- and second-stage least squares estimates in the instrumental variable regressions. Additionally, the recursive demeaning estimator employs the mean as a location metric, which is also sensitive to outliers. PST closely follow Berk and van Binsbergen (2015), who undertake an extensive effort to address data errors in mutual fund databases, and therefore it is not immediately clear whether outliers will be a problem.[6] However, their efforts are largely restricted to reconciling errors in unadjusted fund returns and fund asset sizes. This approach does not consider that the source for an observation's outlyingess could be in fund benchmark-adjusted returns or other fund characteristics.

[6] However, we do not suspect benchmarking problems in Berk and van Binsbergen (2017) as they do not rely on Morningstar categories and instead compute alphas using Vanguard index funds.

PST argue that the Morningstar Category index-adjusted returns more precisely account for style and risk than the commonly used loadings on the Fama-French factors. This approach circumvents the need to address large estimation errors in some mutual fund betas.[7] Cremers et al. (2013) also recommend the use of benchmark-adjusted returns after noting that the Fama-French factor models produce biased performance measurements. Our concern is not with benchmarking per se, rather that the summary-level Morningstar categories were not designed for use in long-run historical analyses but to aid current investment decision-making. As a result, the Morningstar categories used in PST are for the most recent period and do not account for historical variations in fund style. This is a potential problem as funds often vary their actual investment styles over time (e.g., Chan et al. 2002) and many funds do not constrain their managers to hold specific asset types (e.g., Almazan et al. 2004). This means measurement errors in Morningstar Category benchmark-adjusted returns are possible even when the fund returns themselves are correct.[8]

There is also the concern that the Morningstar Category classifications are incorrect. Morningstar does not attempt to create categories that precisely capture every conceivable fund style, but instead creates categories based on the popularity of investment styles. As a result, the number of investment categories in Morningstar can fluctuate significantly. For example, in the middle of 2017, there were 115 Morningstar categories in use but only 49 in December 2000.[9] The implication here is that Morningstar categories do not closely match the investment styles of all funds. As an illustration of the potential for measurement error when using Morningstar categories to compute benchmark-adjusted returns, consider the Mid-Cap Blend Morningstar category. Morningstar assigns funds to this category that hold a mix of mid-cap growth and value stocks and funds that hold a mix of small-cap and large cap-stocks. Obviously, using a single benchmark index for the two very different types of mid-cap designated funds will lead to measurement errors. The potential for substantial performance measurement errors is also exacerbated by the lack of unique Morningstar Category benchmark indices available to compute benchmark-adjusted returns. In the middle of 2017, there were only 89 Morningstar benchmark indices to cover the 115 Morningstar categories.[10]

[7] PST report large standard errors for OLS beta estimates at the 90th, 95th, and 99th percentiles. Our results are robust to using Fama-French three-factor alphas.

[8] See Berk and van Binsbergen (2017) for suggestions on benchmarking mutual fund performance.

[9] From the December 2000 Morningstar Principia installment.

[10] Counts are based on our sample. The 2017 Morningstar guide only lists 110 current investment categories.

2.2 Identifying Potential Data Errors

The first step in addressing outlier and data error concerns is identification. We follow Adams et al. (2018, 2019) and compute outlier-robust standardized residuals as a measure of vertical distance and the robust Mahalanobis distances for horizontal distances.[11] Using the same specification that produces the OLS results in Table 2, Figure 1 displays the plot of the gross market-adjusted return model for the year 2000 with vertical distances on the *y*-axis and horizontal distances on the *x*-axis.[12] The two horizontal lines mark the *y* bounds at ± 2.25, which are values from the standard normal distribution that separate the ± 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 variables space. Figure 1 reports several vertical outlier funds. Funds with large vertical and large horizontal distances are horizontal outliers and located in region H in Figure 1.[13] A horizontal outlier is an observation that is outlying in the independent variable space and located far from the estimated regression line as determined by the overwhelming majority of observations. Horizontal outliers significantly affect the estimation of both the intercept and slope coefficients. Figure 1 shows that many funds in the year 2000 are horizontal outliers.

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 G in Figure 1). It is important to note that these good outliers, like all outliers, can be the result of data errors. The non-outlier funds, which constitute the vast majority of the observations, are located in region N. Overall, Figure 1 shows that the sample contains unusual mutual funds that have the potential to bias coefficient estimates.

[11] Mahalanobis distance is a measure of the multivariate outlyingness of an observation in terms of the explanatory variables, defined as $d_i=\sqrt{(X_i-u)\Sigma^{-1}(X_i-u)'}$, where μ is the multivariate location vector, $\sum$ is the covariance matrix of the explanatory variables, and Xi is the i^{th} row vector of matrix X, for $1 \leq i \leq n$.

[12] We plot a single year only because a figure using all years in our sample would have too many observations to be legible.

[13] Although not reported, Rousseeuw and van Zomeren (1990) plots for the other specifications also show large numbers of vertical and horizontal outlier funds.

2.3 How Different are the Outliers?

Having identified the outliers, we investigate how these errant fund observations vary from typical mutual funds by examining differences in fund characteristics between the two outlier type samples (vertical and horizontal) and the non-outlier funds. Table 3 shows that the differences in the returns are remarkable. The mean and median monthly benchmark-adjusted gross 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 33 basis points per month. The difference in median monthly returns is even larger, approximately 308 basis points or almost 37% per year. The magnitudes indicate that within the vertical- and horizontal-outlier subsamples there are funds with more extreme benchmark-adjusted returns compared to their cohorts. These large differences in benchmark-adjusted returns suggest that outlier observations may influence the estimated coefficients on the relation between scale and fund performance.

We also find statistically significant differences in the independent variables. Non-outlier funds are significantly larger than vertical outlier funds but smaller than the horizontal outlier funds. In terms of industry size, vertical and horizontal outlier funds have smaller mean and median values. This suggests outliers occur in years where the active managed funds industry is smaller, possibly due to lower stock valuations or industry-wide cash outflows from managed funds. We explain this possibility below.

3 Empirical Results

3.1 Evidence on the Nature of Returns to Scale Using PST

3.1.1 The Most Extreme Observations and Their Investment Categories

To examine returns to scale at the fund and industry levels, we regress current month benchmark-adjusted gross fund returns on prior month FUNDSIZE and INDUSTRYSIZE using our primary sample of U.S.-domiciled domestic equity funds for the period March 1993 to December 2014.[14] Table 4 provides the findings. Panel A reports results for the OLS estimates [compare to PST column 7 in Panel A of Table 3], Panel B for the OLS fixed effects estimates [compare to PST column 8 in Panel A of Table 3], and Panel C for the recursive demeaned estimates [compare to PST column 9 in Panel

[14] We also run the same regressions for the period March 1993 to December 2011 to match PST. The results are similar and are provided in Appendix A-2.

A of Table 3]. We quantify the impact of outliers by removing increasing percentages of observations, in the range between 0% and 10%, with large vertical and horizontal distances. For example, for the 1% level we drop observations with vertical and horizontal distances in the top 0.5% of all observations. Because vertical distances can be positive or negative, we also run regressions where observations with the most negative vertical distances are dropped. In those regressions, the estimated coefficients generally have the same sign (negative) and statistical significance as in PST. When observations with the most negative and positive are dropped symmetrically, the estimated coefficients are insignificant. Therefore, we conclude that the coefficient estimates are not sensitive to the observations with negative vertical distances and focus our attention on observations with large positive vertical distances.

Our primary interest is in the results presented in Panel C where we consider both fund- and industry-scale effects using the finite-sample bias-free recursive demeaning approach developed in PST. We note that the results from the full sample, without dropping any observations in Column 1, are close to those reported in PST and in our Table 2 for the sample period March 1993 to December 2011. We also report R^2 values in Table 4 and find they generally rise as the number of outliers removed increases, further confirming that the extreme observations do not fit the pattern of the data.[15]

The remaining columns repeat the tests for increasing percentages of extreme observations removed. The estimated coefficient on FUNDSIZE is insignificant for all dropping levels. The variable INDUSTRYSIZE remains negative but insignificant at the customary 5% level when 0.025% of outlying observations are removed. However, INDUSTRYSIZE becomes insignificant even at the 10% level when the most extreme 0.05% observations are removed. This suggests that between 0.025% and 0.050% of the sample observations are driving the negative returns to scale at the industry-level finding, possibly much closer to 0.025% given the marginal significance. Put differently, for up to 99.975% of the sample we do not find evidence of a negative industry scale effect. About 84 to 168 influential observations, which correspond to the number of observations in the 0.025% and 0.050% quantiles from a sample of 332,516 observations, are driving the negative INDUSTRYSIZE results. To gauge how extreme the 168 influential observations are compared to the full sample, in untabulated

[15] Column 10 reports the number of observations used (299,864) and the number dropped (32,652). The actual percentage of observations dropped is 9.81%, because some funds are in the top 5% in terms of both vertical and horizontal distances. Of the observations removed, about 25% are vertical outliers, 28% are horizontal outliers, and 47% are good outliers. In reported analysis we remove all vertical and horizontal outliers (approximately 34K observations) and find insignificant FUNDSIZE and positively significant (at the 1% level) INDUSTRYSIZE estimated coefficients.

analysis we repeat the recursive demeaning estimation for a sample containing only the 168 influential observations. We find the estimated coefficient on INDUSTRYSIZE is about 288 times larger than for the full sample. For FUNDSIZE the estimated coefficient is positive and five times larger than for the full sample.

Somewhat surprisingly, the estimated coefficient on INDUSTRYSIZE becomes positive and significant when approximately 1% (3,329 observations) of the most extreme outliers are removed. Thus, for 99% of the sample of actively managed equity mutual funds there is an apparent positive scale effect at the industry level. We report similar results in Appendix Table A.2 for the 1993–2011 sample period.

PST argue that OLS fixed effects regressions capture the relation between gross benchmark-adjusted returns and INDUSTRYSIZE better-than-recursive demeaning regressions. We disagree. Because FUNDSIZE and INDUSTRYSIZE are correlated, the finite sample OLS fixed effect bias for FUNDSIZE affects the estimation of INDUSTRYSIZE in the fixed effects specification.[16] Put differently, the recursive demeaning is a better estimator for INDUSTRYSIZE because it minimizes biases in estimates of FUNDSIZE. Nevertheless, in the OLS fixed effects results in Panel B, the INDUSTRYSIZE coefficient becomes insignificant when we drop between 0.25% and 0.50% of the sample observations and becomes positive after removing between 1% and 2.5% (columns 7 and 8) of the most extreme observations.[17] Therefore, while the OLS fixed effects results are somewhat stronger for INDUSTRYSIZE, small numbers of outliers continue to drive the full sample negative relation with gross benchmark-adjusted returns. If decreasing returns to industry scale is a general effect, in that as industry size increases fund performance tends to decrease, it should not be fragile and depend on small numbers of influential observations that are potentially data errors. We also find similar results for the OLS model in Panel A.

In untabulated analysis, we confirm the efficacy of our approach to identifying influential observations by running 1000 OLS, OLS FE, and recursive demeaning regressions and randomly removing 0.05% of the sample observations in each of the 3,000 regressions, the estimated coefficients are very similar to those obtained for the full sample (i.e., column 1 in Table 4). We repeat the regressions by randomly dropping 5.0% of the sample observations and again find results that are

[16] Increased competition leads to new and smaller funds driving down average FUNDSIZE even though INDUSTRYSIZE is increasing.

[17] We find similar OLS fixed effects results when regressing INDUSTRYSIZE on returns without FUNDSIZE in the model.

similar to the full sample results. We conclude that our findings are the result of identifiable influential observations and not simply the result of random combinations of data.

The results in Table 4 are consistent with constant returns to scale at the fund level. However, interpreting the results for any industry-level returns to scale is more problematic. The estimated coefficients on INDUSTRYSIZE are positively significant when 1%–10% (2.5%–10%) of the most extreme observations are removed, insignificant when 0.05%–0.50% (0.50%–1.00%) are removed, and negatively significant when 0.00–0.025% (0%–0.25%) are removed in the recursive demeaning (OLS fixed effect) regressions. However, the correct dropping hinges on the actual level of data errors in the sample. For example, if sufficient data errors are present in the 1% of observations dropped in Column 4 of Panel C, or in the 2.5% of observations in Column 8 of Panel B, we can reject the null hypothesis of constant returns to industry level returns to scale and conclude there are positive industry scale effects. We investigate further the potential for positive economies of scale for mutual funds later in this paper.

3.1.2 Identifying Data Errors

3.1.2.1 The Most Extreme Observations and Their Investment Categories

Given the sensitivity of the estimated coefficients on INDUSTRYSIZE to extreme observations, we next investigate the origin of these influential observations and the reasons for their outlyingness. Of interest is whether the outliers are due to omitted variables, sampling errors, measurement errors, or other errors in the CRSP and Morningstar data. This effort requires multiple resources including Morningstar Direct, prospectuses, annual reports, fund webpages, SEC filings, and news articles. We keep the exercise tractable by focusing on the 0.05% most extreme observations whose inclusion produces the negative returns to industry scale finding in the recursive demeaning regressions.

Panel A of Table 5 lists the top 20 U.S. domestic equity funds that most frequently have vertical and horizontal distances—large returns and size in a multivariate sense—in the top 0.05% or have the largest vertical and horizontal distances. Panel A also includes the Morningstar Benchmark Index for each outlier fund, the number of months the fund among the most extreme 0.05%, and identifies each fund's outlier type (good, vertical, or horizontal).

Of the 262 months in the March 1993 to December 2014 period, the Growth Fund of America from the American Funds tops the list with 84 occurrences, which is a remarkable number of instances and more than random chance can explain. Panel A notes the Morningstar Category of the Growth Fund of America is U.S. Large Growth and its Morningstar Benchmark is the Russell 1000 Growth index. We next review the fund's prospectuses, annual reports, Morningstar Direct, and the fund's webpage for information on why it is so often an extreme observation. The Growth Fund of America is currently the largest actively managed mutual fund offered for sale in the U.S. and has outperformed its Morningstar Benchmark Index by an average of 3.5% per year over our sample period. This explains why the fund is so often a good outlier, one that falls close to the regression line formed by the bulk of the other observations, thereby reducing standard error estimates. It is a large fund with positive benchmark-adjusted returns and the OLS estimated coefficient on fund size for the bulk of the data is positive. That is, the fund's size and performance are consistent with the OLS results in Panel A of Table 4 for 1%, 2.5%, 5%, and 10% of outliers removed.

However, the Morningstar-assigned Investment Categories and Benchmark Indices do not appear to capture the Growth Fund of America's investment objectives and practices. A review of the fund's holdings indicates it is not a typical domestic large capitalization growth fund. For example, the fund's 2001 N-30D filing shows over 25% of assets are invested in international stocks and short-term securities. The fund's holdings consistently include small- and middle-capitalization common stocks, preferred stocks, convertibles, international stocks, and fixed income securities. These holdings are consistent with the fund's flexible investment strategy where its managers face few restrictions on the types of assets it holds. It must hold at least 65% of assets in common stocks and may invest up to 25% (up from 15% in 2010) of its assets in securities domiciled outside the U.S.

We also note that in most years only about half of its holdings are large-capitalization growth stocks as defined by Morningstar. It is also worth noting that the fund's unusually long-tenured managers do not classify it as a typical growth fund that invests in growth stocks as defined by Morningstar; rather, they focus on capital appreciation regardless of asset type. The Growth Fund of America's 2014 Annual Report states, "The best companies, the best investments—those can be pretty subjective. Thankfully, the focus and structure behind GFA allows portfolio managers and analysts to determine "best" as they see fit—and they're not constrained by arbitrary definitions. There are growth funds out there that, when they say growth, they mean they invest in 'growth' stocks as determined by Morningstar or Russell, and it is often code for simply high valuation, which is not necessarily good for shareholder return, … We're about capital appreciation, flexibly seeking the best opportunities no matter how they appear. Often, it is a traditional growth company, sometimes it is a turnaround or a down-and-out. We gravitate toward whatever we think is most attractive that fits our objective of capital appreciation." Overall, the evidence indicates that the Growth Fund of America's performance is poorly measured by its Morningstar benchmark-adjusted gross returns. It also seems reasonable to question whether the fund should be included in the sample as the construction process of PST explicitly calls for the removal of non-domestic and non-equity funds. Because Growth Fund of America is a good outlier, including or excluding it only minimally affects coefficient point estimates.

The second most frequently occurring fund listed in Panel A is the Comstock Capital Value Fund, which has a U.S. Fund Bear Market Morningstar Category classification and an S&P 500 Morningstar Benchmark Index. The Comstock Capital Value Fund appears in the top 0.05% most extreme of all observations seven times and is a horizontal outlier. Bear market funds take short positions and attract investors seeking to profit from market downturns. The fund's principal investment strategies allow its managers to invest in a wide range of asset classes and market sectors. Investments may include domestic and foreign equity and debt securities, money market instruments, and derivatives. The fund is not managed as a diversified portfolio, may take either long or short positions, and shifts investments frequently. Reviewing the fund's holdings history, we find the fund often holds long positions in U.S. Treasuries and short positions in index futures contracts and equities. We examine the holdings of other bear market funds and find similar portfolio holdings patterns. Table 5 lists three additional bear market funds: the Federated Prudent Bear Fund, the Grizzly Short Fund, and the Comstock Strategy Fund. The inclusion of these funds is a sampling error.

The next two funds listed in Panel A are the Matthews Korea Fund and the Voya Russia Fund. Both are U.S. Fund Miscellaneous Region funds benchmarked against the MSCI ACWI Ex USA index, horizontal outliers, and in the top 0.05% of most extreme observations four times. PST use keyword searches of Morningstar Category names to identify and remove funds from 15 of 16 international equity fund categories from their sample. We suspect not excluding funds from the remaining Miscellaneous Region Morningstar Category is an oversight.

The seventh fund in the list of most extreme outliers is the Nuveen Mid Cap Select Fund, formerly the First American Funds Mid-Cap Select Fund, which was formerly the First American Funds Small-Mid Core Fund, which was formerly the First American Funds Technology Fund. First, American Funds was purchased by Nuveen in 2010 and the fund changed its investment objective from investing in technology stocks to stocks of small-cap and mid-cap companies in 2005 and from small-cap and mid-cap companies to mid-cap companies in 2009. The firm is classified as a Mid-Cap Blend fund in the current Morningstar database but the numerous name changes point to substantial shifts in investment strategy that likely make the Morningstar-assigned benchmark index a poor match in some months. We confirm this by reviewing the Morningstar Equity Style Box data, a 3x3 grid of investment style (value, blend, growth) and stock capitalization (small, mid, and large), which are historically available. At different points during the sample period the fund is classified as a small-cap growth, a mid-cap growth, a large cap growth, and a mid-cap blend. The Nuveen Mid Cap Fund is a most extreme horizontal outlier three times during the sample period. Three factors contribute to its outlyingness. First, Chen et al. (2004) document returns to scale at the family level. Because the Mid Cap Select fund was acquired by a new fund family, this may present a possible omitted variable problem. Second, Cooper et al. (2005) find that fund name changes affect fund flows. The numerous name changes potentially affect both return and size and thereby could represent another omitted variable problem. The third factor is simple measurement error. Using a single benchmark index that does not consistently match the underlying portfolio changes causes data errors. Overall, it is prudent to remove this fund from the sample.

The eighth and eleventh most extreme outlier funds, the Embarcadero Absolute Return and Market Neutral funds, also experienced family name (formerly Van Wagoner Capital Management), fund name, and investment objective changes. These non-equity funds were not dropped by PST filters and are a mistake in Morningstar's Benchmark classification scheme. In addition, Van Wagoner settled civil fraud charges brought by the SEC over claims the family misstated the value of illiquid securities in the funds.

The remaining funds in Panel A include bear market and long-short funds, international funds, funds that experienced name and investment objective changes during the sample period and funds whose investment objective is not well captured by the Morningstar Benchmark indices used to compute the Morningstar Category benchmark returns. In addition, the remaining 0.05% of funds not listed in Panel A also represents data errors. Thus, all of the influential observations that drive the negative returns to industry scale found in Panel C of Table 4 are examples of data errors.

Based upon our finding, we create a revised sample after removing all observations with Morningstar categories of Bear Market, Miscellaneous Region, and Long-Short Equity as well as funds with the names of Market Neutral, Internet, IPO, and Russia. That is, we create a sample after removing all obvious sampling errors identified by the most extreme 0.05% list in Panel A and rerun the recursive demeaning regressions in Panel C of Table 4. The new results indicate that the number of influential observations falls from 168 to only 69, all of which are either benchmarking or omitted variable errors. Thus, data errors representing only 0.02% (69 observations) of the revised sample drive the diseconomies of industry scale result.

3.1.2.2 Cleaning the Data

Our results suggest that PST's main findings are driven by a small number of data errors despite their data construction efforts. In this section, we apply a multi-stage cleaning process to remove errors and examine how their removal affects the sample and results. PST argue that the relation between gross benchmark-adjusted returns and industry scale is better captured by OLS fixed effects. While we have argued that recursive demeaning is the better estimator, nevertheless we next examine how many of the 1,672 observations (0.50% of the sample) that drive the OLS fixed effects results for INDUSTRYSIZE are data errors.

Panel B of Table 5 provides the findings from OLS fixed effects regressions after cleaning the data in three stages. First, the Morningstar categories used in PST are for the most recent period and do not account for historical variations in fund style. Therefore, we remove observations where the current Morningstar category assignment differs from the historical category assignment (i.e., Morningstar's actual category assignment at each point in time). We classify these as Current vs. Historical Category errors. The OLS fixed effects point estimate on INDUSTRYSIZE falls to –0.013 (t-stat of 2.12) in Column 1 from the full sample estimate of –0.022 (t-stat of 3.28) as reported in Table 4, Panel B. The recursive demeaning FUNDSIZE and INDUSTRYSIZE slope estimates in Column 2 are economically and statistically insignificant. Panel B reports the 679 Current vs. Historical Category errors account for 0.20% of the sample and 40.61% of the 1,672 observations that drive the negative industry scale result.

Second, we identify observations where the most recent Morningstar category assignment differs from the historical style assignment. Morningstar classifies a fund's holdings style (e.g., Large Cap Growth, Small Cap Value, etc.) for each reporting period. Morningstar then uses three years of style classifications to assign a fund to a category. We classify instances where a fund's most recent Morningstar Category assignment differs from the historical style assignment as Current vs. Historical Style errors. Columns 3 and 4 report an additional 74 style data errors (most style errors are also category errors).

Third, we manually identify 289 additional data errors. PST use Morningstar Category name keyword searches for "bond funds, money market funds, international funds, funds of funds, industry funds, real estate funds, target retirement funds, and other non-equity funds" to remove sampling and benchmarking data errors. Instead, we use category—and fund name—keyword searches to identify suspect observations. We also use the frequency of occurrence to search for confounding events including family and fund M&As, SEC investigations, name changes, and other events that potentially affect investor flows and fund size. We then classify data errors using information from fund webpages, SEC filings, and news articles.

Our data cleaning process accounts for data errors comprising 0.31% of the sample and 62.32% of the most extreme 1,672 observations. The data errors tend to be smaller funds (mean TNA of about $50 million) and have large variations in monthly gross benchmark adjusted returns (-2% to +42%). More importantly, both the OLS fixed effects and recursive demeaning INDUSTRYSIZE coefficient estimates are insignificant after removing the data errors.[18] In sum, Table 5 demonstrate the importance of robust data cleaning strategies to mitigate benchmark-adjusted return measurement, omitted variable, and sampling errors.

3.1.3 Incidence of Most Extreme Data Errors by Year

Combined, Tables 4 and 5 demonstrate that a few data errors are responsible for the diseconomies of industry-scale result. During our manual verifications, we find the errors cluster around the years 2000 and 2008. This suggests data errors are more likely in periods of increased market volatility.

Table 6 provides the distribution of influential observations by year for our main sample period. The first column lists the year while the second and third columns list the number of top 0.05% and 0.50% most extreme observations occurring in each year. Table 6 reports that the incidence of influential data errors varies considerably by year. There are no top 0.05% influential errors in the periods 1993–1997, 2002–2003, and 2013–2014. For the remaining years, there are two clusters of influential data errors that correspond to periods of stock market price instability and net investor cash outflows from mutual funds. The first cluster from 1999 to 2001 occurs around the market crash following the Internet bubble in March 2000. The second cluster occurs around the 2008 financial crisis. The distribution of influential observations in the first cluster is highly concentrated with most occurring in 2000. In fact, the number of influential observations occurring in 2000 is more than double that of 2008. While influential observations occur with less frequency in the second cluster, they take place over a longer period, perhaps reflecting the nature of the slow development of the financial crisis and the subsequent recession. The top 0.50% influential observations follow the same general pattern as the top 0.05% influential errors. Figure 2 shows the 10^{th}, 50^{th}, and 90^{th} percentiles of monthly benchmark-adjusted returns. Consistent with the annual counts of data errors reported in Table 6, Figure 2 reports more extreme benchmark-adjusted returns around the 2000 and 2008

[18] Our suspect fund approach likely undercounts the actual number of data errors. As a result, the point estimates represent an upper bound.

clusters. This fits with our intuition that data errors, especially benchmarking errors, are more likely in periods of increased market volatility.

3.1.4 Regressions of Fund Performance on Fund Size and Industry Size in the Post-2001 Period

It is worth noting that all but seven of the 168 influential data errors in the second column of Table 6 occur during PST's main sample period of 1993–2011. This indicates that the influential data errors are not the result of our extended sample period where we add an additional three years of fund data, 2012–2014. This also suggests that the diseconomies-of-scale results are sample time–specific due to clustering of data errors. To check, we rerun the OLS fixed effects and recursive demeaning regressions of fund gross benchmark-adjusted returns on prior month FUNDSIZE and INDUSTRYSIZE but for the January 2002 through December 2014 period. We choose the start date of January 2002 to exclude the dot.com crash as well as the market shock following September 11, 2001. Table 7 reports the findings for OLS fixed effects in columns 1 and 2 and for recursive demeaning in columns 3 and 4. Columns 1 and 3 report coefficient estimates when no observations are removed and columns 2 and 4 provide slope coefficients and *t*-statistics after the most extreme outliers are removed.[19] The results for FUNDSIZE and INDUSTRYSIZE are insignificant for the full sample and when the most extreme observations are dropped.[20] That INDUSTRYSIZE is insignificant whether or not the influential observations are removed suggests the PST findings are due to data errors clustered around the dot.com stock market crash of 2000.

3.1.5 Using Fama-French Factor-Adjusted Returns

Our manual investigations reveal that all of the 0.05% (168 observations) recursive demeaning regression influential observations are either omitted variable errors, sampling errors, or benchmarking errors where incorrect or imprecise indexes are used to compute benchmark-adjusted returns. Likewise, in the OLS fixed effects regressions our tests indicate that less than 0.50% (1,682) of the sample observations are influential. We examine these 1,682 influential observations and find several incidences of sampling errors including Bear Market, Market Neutral, Specialized Sector, and International funds. However, manually examining the historical holdings and other fund information for omitted variable and benchmarking errors for all funds is a much larger task. Instead, we employ

[19] We also examine the OLS and OLS FE models and find INDUSTRYSIZE is insignificant.

[20] In unreported regressions, we find insignificant FUNDSIZE coefficient at all outlier removal percentages. We also find positive and significant INDUSTRYSIZE coefficients when 5% of extreme outliers are dropped from the sample.

Fama-French three-factor alphas to avoid measurement errors in Morningstar Category benchmark-adjusted returns.

Table 8 presents the results when removing increasing percentages of extreme observations. Panels A and B present the OLS fixed effects results for the full sample and after removing sampling errors, respectively, while Panel C provides the recursive demeaning regression results. In Panel A, we find the OLS fixed effects estimated coefficient estimate on INDUSTRYSIZE to be negative and insignificant at the 5% level. It is also insignificant when we remove only 0.025% (84 of 334,271 observations) of the most extreme observations. As with the Morningstar benchmark-adjusted returns, INDUSTRYSIZE becomes positive and significant when dropping about 5.00% of the sample. We also find in Panel B that after removing bear market, market neutral, specialized sector, and international funds (i.e., sampling errors) the estimated coefficient on INDUSTRYSIZE is insignificant with the removal of 0.00–5.00% extreme observations and positively significant thereafter. Lastly, the recursive demeaning estimated coefficients on FUNDSIZE and INDUSTRYSIZE are insignificant for the full sample and all dropping rates.

The earlier results from regressing Morningstar Category benchmark-adjusted returns are sensitive to removing small numbers of influential observations. Subsequent manual examinations against outside sources reveal all of the recursive demeaning influential observations are sampling, omitted, or benchmark-adjusted return errors clustered around the dot.com stock market crash of 2000. We find an insignificant relation between benchmark-adjusted returns and industry scale in the post-2000 period and conclude data errors are responsible for the industry size results. For robustness, we employ Fama-French adjusted returns in lieu of Morningstar Category benchmark-adjusted returns and fail to reject the null hypothesis of constant returns to scale at both the fund and industry levels. However, PST argue that OLS fixed effects regressions better capture the relation between returns and industry size. Even so, we find an insignificant estimated coefficient for industry size. These findings are consistent with our contention that data errors are responsible for the full sample Morningstar Category index-adjusted return results.

3.2 Evidence on the Nature of Returns to Scale Using Zhu (2018)

In this section, we investigate whether our fund level constant returns to scale are robust to Zhu's (2018) recursive demeaning approach. Zhu (2018) reexamines the PST study using an enhanced empirical strategy by adding an intercept to the first-stage regression. She argues that the method of PST suffers an inherent misspecification resulting from a model restriction, which is problematic for the fund size process. Unlike PST, Zhu (2018) shows that diseconomies of scale exist at the fund level. She includes only funds that fall into one of the nine size (small, mid, and large capitalization stocks) and style (value, blend, and growth) intersected Morningstar categories to avoid bond, international, sector, money market, and other non-equity funds. However, the implications of this data restriction on the robustness of the empirical results are unclear. While reducing the number of investment categories relative to PST yields a reduction in potential misclassifications, using less size-style categories provides less precision because only nine indexes are used to compute benchmark-adjusted returns and there are considerable variations in investment policies and portfolio holdings within each category.

3.2.1 Empirical Results

We follow Zhu's (2018) sample generation method carefully with the exception that we begin with the merged CRSP-Morningstar database that reconciles errors in fund returns and fund asset sizes rather than the raw Morningstar database. To maintain consistency with our earlier findings, our sample begins in March 1993 whereas Zhu (2018) uses a January 1995 start date. Again, it remains unclear how our sample selection criteria will affect the results. If diseconomies of scale are economically important, they will be resistant to slight changes in sample construction. On the other hand, small changes in the sample could matter if the documented diseconomies of returns to scale are economically trivial or driven by bad data and measurement error–induced outliers.

Table 9 presents our results when removing increasing percentages of extreme observations from the sample. The results are provided for two measures: FUNDSIZE (or the dollars of assets under management) in Panel A, and the natural log of FUNDSIZE in Panel B. Zhu (2018) argues the natural log of FUNDSIZE is a better measure because of severe (positive) skewness in dollar FUNDSIZE. However, for the recursively-demeaned log of FUNDSIZE, skewness is non-trivially negative. That is, the distribution of the population for the log of FUNDSIZE and the log of FUNDSIZE measure used in the recursive demeaning regression tests are different. Because our focus is on the data and not on the empirical method, we do not address further the issue of which functional form of fund size is most appropriate.

Panel A of Table 9 reports insignificant results at the dollar fund size level regardless of how many observations are removed. This result differs from the negatively significant estimated coefficient on dollar FUNDSIZE found in Zhu (2018). Instead, the insignificant findings are consistent with PST as well as those in this study. In unreported regressions, we find a statistically significant negative dollar

FUNDSIZE coefficient when the sample starts in January 1995 as in Zhu (2018) rather than our March 1993 start date. However, when removing only 0.05% of the most extreme observations the estimated coefficient on dollar FUNDSIZE is again insignificant. The fragility of dollar FUNDSIZE coefficient estimates to slight variations in sample counts and periods is indicative of a sample-specific relation and is therefore not generalizable. In addition, Table 9 reports positive economies to industry scale when the top 1% of most influential observations are dropped, evidence consistent with our earlier results in Table 4.

Panel B of Table 9 reports negative and statistically significant results for log of FUNDSIZE until removing between 2.5% and 5.0% of the sample. In terms of economic significance, the estimated coefficient is economically small at -7 basis points when dropping 2.5% of the sample and zero when dropping 5% or more. In subsequent iterative regressions, we narrow the number of influential observations to 4.0% of the overall sample or 11,996 fund-month observations from 1,379 funds. That is, removing the most multivariate extreme 4% of observations causes the estimated coefficient on log of FUNDSIZE to become statistically insignificant.

3.2.2 A Closer Look at the Zhu (2018) Data

When reviewing the 20 largest and most frequently occurring influential funds, we see considerable overlap with funds that experienced multiple name, investment objective, and Morningstar Category changes listed in Table 5. The Embarcadero Absolute Return and Market Neutral funds (non-equity portfolios and SEC fraud charges) and Nuveen funds are examples of funds that are on the top 20 list of both samples. The top 20 influential funds (for Panel B of Table 9) are mainly unconstrained, multi-capitalization, and multi-style funds. For example, the Midas Magic Fund, a Large Cap fund, may invest in any security type (common and preferred stocks, bonds, convertibles, etc.), in any industry sector, in domestic or foreign markets, and in companies of any size. Another top 20 influential fund is the Fidelity Select Construction and Housing Portfolio, classified by Morningstar as a Mid-Cap Growth fund.

Influential funds outside the top 20 include the MFS Managed Sectors fund (implicated in 2002–2003 late trading scandal), micro-cap funds, poorly matched mid-cap funds (see discussion in section 2.1 on performance measurement error issues in mid-cap classified funds), and contrarian funds (low correlation with benchmark index). Other influential funds experienced family-level mergers and acquisitions, have names in Morningstar that include the words ‘do not use,’ total return funds, growth and income funds, leveraged bull funds, income funds, global IPO funds, multinational funds, and

Internet sector funds. In summary, influential funds tend to have investment strategies not well captured by the nine Morningstar intersected capitalization and style categories.

We compile two-way quintile sorts of gross benchmark-adjusted returns and lagged fund size across the typical and influential fund segmentations. Table 10 reports mean and median gross benchmark-adjusted returns and the number of observations for each of the 25 size and return intersections. Results for the typical fund observations (96% of the sample) are on the left and the influential fund observations (4% of the sample) are on the right. The last row reports the differences between the largest and smallest quintiles. When performance is low, i.e., in the lowest return quintile, largest quintile funds perform less poorly than smallest quintile. In contrast, for the highest quintile of returns, small funds outperform big funds on average. The differences are greater for the influential fund observations. Table 10 shows that most influential observations occur in the highest return quintile. For the influential funds, the performance range of the smallest quintile funds is about 20% per month across the five return quintiles while only about 10% per month in the largest quintile. Overall, the results in Table 10 suggest any relation between returns and scale is primarily in the potential data errors and varies by portfolio size.

In sub-sample analysis, Table 11 presents results from regressing gross benchmark-adjusted returns on *ln*FUNDSIZE and INDUSTRYSIZE by fund size quintile. Panel A of Table 11 reports values for the overall sample that includes typical and influential fund observations. The estimated coefficients for *ln*FUNDSIZE are negative and significant for quintiles one and two, positive and significant for quintiles three and four, and insignificantly different from zero for the largest quintile. Panel A also presents regression results for the combined first and second quintiles in column 6 and the combined third and fourth quintiles in column 7. The *ln*FUNDSIZE estimated coefficients in columns 6 and 7 are essentially identical except for sign. Thus, Panel A presents evidence for diseconomies of scale for small funds, positive economies of scale for intermediate size funds, and constant returns to scale for large funds. INDUSTRYSIZE is positive and significant for quintile 2, negative and significant for quintile 3, but insignificant elsewhere. It is difficult to reconcile theory with these results.

However, by this point we have demonstrated that regressions with data errors can produce fragile results. Panel B of Table 11 provides outlier sensitivity tests for the three apparent scale results. Diseconomies of scale (combined quintiles 1 and 2 results) cease to be statistically significant when dropping more than 2.5% of the sample. However, after dropping 2.5%, the estimated coefficient on *ln*FUNDSIZE is economically small. Curiously, at the 10% dropping level *ln*FUNDSIZE is positive and statistically significant at the 1% level. For the apparent positive returns to scale results (quintiles 3 and 4), dropping more than 1% of the most extreme observations causes the *ln*FUNDSIZE coefficient to be insignificant and after dropping 10% it becomes negative and significant. For apparent constant returns to scale (largest quintile), the estimated coefficient on *ln*FUNDSIZE is insignificant when up to 2.5% dropping and negatively significant beyond 5% dropping.

3.2.3 Economies of Scale at the Fund Level

The analysis in this section, using Zhu's (2018) recursive demeaning estimator, demonstrates the fragility of fund-level economies of scale. The dollar FUNDSIZE measure of scale is significant only for specific sample periods and even then removing small numbers of observations causes it to become insignificant. The results for log of FUNDSIZE are more problematic. Using the entire sample requires us to ignore potential data errors and argue that scale is a tail-risk phenomenon and not a general effect. However, we would then need to explain diseconomies for small funds, positive economies for intermediate-sized funds, and constant returns for large funds. Similar inconsistencies exist for all of the outlier dropping levels. For example, if we assume most of the full sample influential observations are data errors and drop at the 2.5% level, the *ln*FUNDSIZE estimated coefficient is negative for the full sample but not for most funds (i.e., insignificant *ln*FUNDSIZE results for quintiles 3, 4, and 5). However, even then we would need to recognize the negative scale effect is trivially small. The results for industry-level effects are more consistent in that none of the tests shows support for diseconomies of scale and even modest dropping around the 1% to 2.5% level provides evidence for positive economies of scale.

The main takeaway from our analysis of apparent diseconomies of scale is that recognizing data problems is fundamental to examining the economies of scale in the mutual fund industry. After considering errors in the data and the fragility of the results, we fail to reject the null hypothesis of constant returns to fund-level diseconomies of scale. At the industry level, the evidence strongly supports non-negative returns to scale and there is evidence of positive economies of scale.

4 Non-Negative Economies-of-Scale Puzzle: Evidence from Index Funds Only

In Berk and Green's (2004) equilibrium framework, a manager's ability to generate positive gross alpha decreases as fund size increases because her positive NPV investment ideas are not infinitely scalable. Constant or positive returns to fund scale imply an infinite supply of positive NPV investment opportunities, which is clearly unrealistic as the fund would grow infinitely large and become the entire market. PST provide empirical support for diseconomies of scale at the industry level and constant returns to scale at the fund level. These results support the Pástor and Stambaugh (2012) explanation of why the actively managed investment industry is so large given its poor performance relative to passive benchmarks. The assumption of decreasing returns to scale is also central to their model. As money flows into the universe of actively managed funds it becomes more difficult for individual funds to exploit asset mispricing and thereby outperform passive benchmarks, resulting in decreasing returns to scale. As investors learn about active-managed funds' underperformance they reduce their allocation, which in turn leads to more mispricing and this implies higher future performance. Therefore, investors divest less than they would if returns to scale were constant. In the model of Pástor and Stambaugh (2012) the active management industry would not exist if returns are constant and unrelated to industry size as rational investors would perceive negative alpha at any industry size and would not allocate any monies to active management.

We attempt to reconcile our findings of constant returns to fund scale and constant or even positive returns to industry scale by taking a closer look at the Berk and Green (2004) separation of fund investment policy into active and passive components. Berk and Green argue that fund managers act to maximize total fees but must provide expected returns that are at least what investors could obtain by passively investing in index funds in order to attract investor cash inflows. These managers face increasing marginal costs (e.g., additional resources to identify mispriced securities) on monies they actively manage, so there is an optimal amount to be actively managed and they invest the excess in passive benchmarks. Thus, total assets under management include actively and passively managed components. Berk and Green (2004) assume the per-unit operational costs are fixed. Operational costs comprise not only the management, administrative, marketing, and other fees that are included in fund expense ratios, but also the trading commissions charged by brokers, securities lending fees, bid-ask spreads, and price impacts from large trades that influence fund returns. While bid-ask spreads and price impacts from large trades are typically associated with the actively managed portion of fund assets, the passively managed portion also experiences these costs to some degree.

Elton et al. (2012) find positive economies of scale in actively managed funds in that expense ratios decrease as funds get larger. Larger funds may also be able to leverage their scale to obtain lower trading commissions, higher securities lending income, and to efficiently manage trades to minimize bid-ask spreads and trade-related price impacts. The empirical evidence for non-expense ratio operational economies is mixed. Using mutual fund trade data and estimates of transaction costs, Edelen et al. (2007) find trading costs are a major source of diseconomies of scale. Alternatively, using a sample of Canadian funds that are required to report their trades and a measure that captures both transaction costs and trade-related price impact, Christoffersen et al. (2006) find positive scale economies. Evans et al. (2017) report a positive relation between fund size and the likelihood of reporting securities lending income and Adams et al. (2014) find larger funds obtain higher returns on lent securities.

However, none of these studies directly examines how the sum of all non-expense-ratio operating costs is related to size. If net-per-unit non-expense-ratio operating costs are not fixed but decreasing in fund size, the observed empirical relation between gross benchmark-adjusted returns and fund size can be negative, constant, or positive depending on which effect dominates: decreasing returns to scale for actively managed monies or increasing returns to scale for operating costs. Fortunately, we can directly test the hypothesis of positive non-expense-ratio economies to scale in passively managed funds using a sample of index funds. The gross benchmark-adjusted return of an index fund neatly captures the sum of trading commissions, securities lending fees, bid-ask spreads, and price impacts from large trades that affect fund returns (e.g., Harris and Gurel, 1986).

Table 12 reports the slope coefficients and *t*-statistics of FUNDSIZE and INDUSTRYSIZE where the dependent variable is index fund gross benchmark-adjusted return. The index fund sample is constructed using the same process that generated our main sample used in Table 4 with the exception that actively managed funds are removed and index funds are retained. FUNDSIZE is computed as before and INDUSTRYSIZE is the total net assets of all passively managed mutual funds divided by the market capitalization of all CRSP stocks, which is analogous to how INDUSTRYSIZE is computed for actively managed funds. As in Table 4, we drop increasing percentage of observations with the largest vertical and horizontal distances.

The results in Table 12 provide evidence for positive economies of scale in operational performance. The estimated coefficients for both FUNDSIZE and INDUSTRYSIZE are positive and significant when 2.50% of the sample observations are dropped. As before, we manually examine each of the 621 (out of 25,256 fund-month) observations whose inclusion drives the full-sample results and find sampling and benchmarking errors. Our finding of positive economies of scale in index funds and the index fund industry provides context for the apparent puzzle in the actively managed results. That is, even if there are diseconomies in the actively managed portion of a portfolio there could also be positive economies for the passively managed portion. If so, net result would be constant or positive fund-level economies of scale that extends to the industry level.

5 Concluding Remarks

Our findings of non-negative returns to scale appear at odds with Berk and Green's (2004) rational expectations framework and the Pástor and Stambaugh (2012) view that the actively managed mutual fund industry cannot exist under non-negative returns to scale. However, the purpose of this or any replication is not to confirm or reject theory. Rather, replications assess the reliability of empirical efforts. Our failure to find diseconomies of scale does not mean that size is unimportant, only that there is insufficient evidence in PST and Zhu (2017) to empirically reject the null hypothesis of constant returns to scale.

Table 13 shows considerable variations in the empirical returns to scale literature. Our analysis shows that important empirical diseconomies of scale findings are artifacts of identifiable data errors and extreme outliers. When removed, we cannot reject the hypothesis of constant returns to active fund and industry size. Because the actively managed investment industry has not disappeared and returns to scale cannot empirically be shown to be negative, the puzzle of why the industry is so large given its performance remains. One possible explanation is that although our analysis is robust with respect to data and methodology, we incorrectly retain the null hypothesis of constant returns to scale because of an omitted variable problem. We offer one possible explanation, that even if there are diseconomies arising from a limited supply of positive NPV investment opportunities, there could also be positive economies in non-expense ratio operating costs.

Alternatively, are constant returns to scale simply a matter of reduced asset mispricing investment opportunities being offset by increases in fund manager skill? That is, as the active management industry becomes more competitive, low-skill managers and their associated underperformance are displaced by high-skill managers via investor flows. If so, high-skill managers may not find many good investment opportunities (i.e., less asset mispricing) but they avoid bad investments. In which case, negative returns to scale exist in the sense that there are fewer asset mispricing opportunities, but increases in manager skill masks the effect, which makes identifying negative returns to scale empirically difficult.

Another explanation lies in Pástor and Stambaugh's (2012) assumption of available passive benchmarks that are sufficient for pricing assets in an efficient market. Otherwise, mispricing is not reduced by investor flows and there are no diseconomies of scale. Put differently, investors need reliable benchmarks to assess fund performance and make allocation decisions. Because not all active managers follow investment strategies that are neatly captured by investment category classifications and hold assets that are well matched by available passive benchmarks, it may be difficult in practice for investors to identify mispricing and assess fund manager skill. Pástor and Stambaugh (2012) note this difficulty and suggest further empirical work is warranted. We concur and also suggest more work in developing reliable performance benchmarks is likewise warranted.

We conclude by noting that financial research often involves big data and there is a tendency to discount or fail to consider the potential for a few bad data points to bias statistical inferences. Our finding that less than 0.05% of a very large sample biases regression coefficient estimates highlights the importance of starting with good data regardless of sample size. PST employ an algorithmic approach that is effective in eliminating errors in fund sizes and raw returns. Nevertheless, the merged CRSP-Morningstar database still contains data errors so it is apparent that while an algorithmic approach is necessary, more work is needed. Our approach of considering unusual combinations of fund and industry size with returns identifies influential observations, and when combined with algorithmic approaches, makes manual examinations of the data feasible.

References

Adams, J. C., D. Hayunga, and S. A. Mansi. 2018. "Diseconomies Of Scale In The Actively-Managed Mutual Fund Industry: What Do The Outliers In The Data Tell Us?". *Critical Finance Review*. 7(2): 273–329.

Adams, J., D. Hayunga, S. A. Mansi, D. M. Reeb, and V. Verardi. 2019. "Identifying And Treating Outliers In Finance". *Financial Management*. 48(2): 345-384.

Adams, J. C., S. A. Mansi, and T. Nishikawa. 2010. "Internal Governance Mechanisms And Operational Performance: Evidence From Index Mutual Funds". *Review of Financial Studies*. 23(3):1261-1286.

Adams, J. C., S. A. Mansi, and T. Nishikawa. 2010. "Affiliated Agents, Boards Of Directors, And Mutual Fund Securities Lending Returns". *Journal of Financial Research*. 37(4): 461–494.

Almazan, A., K. C. Brown, M. Carlson, and D. A. Chapman. 2004. "Why Constrain Your Mutual Fund Manager?" *Journal of Financial Economics*. 73(2): 289–321.

Berk, J. B., and R. C. Green. 2004. "Mutual Fund Flows And Performance In Rational Markets." *Journal of Political Economy*. 112(6): 269–1295.

Berk, J. B. and J. H. van Binsbergen. 2015. "Measuring Skill In The Mutual Fund Industry". *Journal of Financial Economics*. 118: 1–20.

Berk, J. B. and J. H. van Binsbergen. 2017. "Mutual Funds In Equilibrium". *Annual Review of Financial Economics*. 9: 147–167.

Bhojraj, S., Y. Cho, N. Yehuda. 2012. "Mutual Fund Family Size And Mutual Fund Performance: The Role of Regulatory Changes". *Journal of Accounting Research*. 50(3):647–684.

Busse, J., L. Jiang, and Y. Tang. 2021. "Double-Adjusted Mutual Fund Performance". *The Review of Asset Pricing Studies*. 11(1): 169-208.

Chan, L. K. C., H. Chen, and J. Lakonishok. 2002. "On Mutual Fund Investment Styles". *Review of Financial Studies*. 15: 1407–1437.

Chen, J., H. Hong, M. Huang, and J. D. Kubik. 2004. "Does Fund Size Erode Mutual Fund Performance? The Role Of Liquidity And Organization". *American Economic Review*. 94: 1276–1302.

Cohen-Freue, G. V., H. Ortiz-Molina, and R. H. Zamar. 2013. "A Natural Robustification Of The Ordinary Instrumental Variables Estimator". *Biometrics*. 69(3): 641–650.

Cooper, M. J., H. Gulen, and P. R. Rau. 2005. "Changing Names With Style: Mutual Fund Name Changes And Their Effects On Fund Flows". *Journal of Finance*. 60: 2825–2858.

Cremers, M., Petajisto, A., and E. Zitzewitz. 2013. "Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation". *Critical Finance Review*. 2: 1–48.

Christoffersen, S. K., D. B. Keim, and D. K. Musto. 2006. "Valuable Information And Costly Liquidity: Evidence From Individual Mutual Fund Trades". University of Toronto Working Paper.

Crane, A. D. and K. Crotty. 2018. "Passive Versus Active Fund Performance: Do Index Funds Have Skill?" *Journal of Financial and Quantitative Analysis*. 53(1): 33–64.

Edelen, R. M., R. B. Evans, and G. B. Kadlec. 2007. "Scale Effects In Mutual Fund Performance: The Role of Trading Costs." University of California, Davis Working Paper.

Elton, E. J., M. J. Gruber, and C. R. Blake. 2012. "Does Mutual Fund Size Matter? The Relationship Between Size And Performance". *Review of Asset Pricing Studies*. 2: 31–55.

Evans, R., M. A. Ferreira, and M. P. Prado. 2017. "Fund Performance And Equity Lending: Why Lend What You Can Sell?" *Review of Finance*. 21(3): 1093–1121.

Fama, E. F. and K. R. French. 1996. "Multifactor Explanations Of Asset Pricing Anomalies". *Journal of Finance*. 51: 55–87.

Ferreira, M. A., A. Keswani, A. F. Miguel, and S. B. Ramos. 2013. "The Determinants Of Mutual Fund Performance: A Cross-Country Study". *Review of Finance* 2: 483–525.

Golez, B. and S. Shive. 2015. "When Fund Flows Take The Fun (Alpha) Away". University of Notre Dame Working Paper.

Harvey, C. and Y. Liu. 2017. "Decreasing Returns To Scale, Fund Flows, And Performance". Duke University Working Paper.

Harris, L. and E. Gurel. 1986. "Price And Volume Effects Associated With Changes In The S&P 500 List: New Evidence For The Existence Of Price Pressures". *Journal of Finance*. 41(4): 815–829.

Hong, H., W. Jiang. 2018. "Diseconomies Of Scale In The Actively-Managed Mutual Fund Industry: What Do The Outliers In The Data Tell Us?: A Response". *Critical Finance Review*. 7(2): 373–377.

Pástor, L., and R. F. Stambaugh. 2012. "On The Size Of The Active Management Industry". *Journal of Political Economy*. 120(4): 740–781.

Pástor, L., R. F. Stambaugh, and L. A. Taylor. 2015. "Scale And Skill In Active Management". *Journal of Financial Economics*. 116, 23–45.

Pollet, J., M. Wilson. 2008. "How Does Size Affect Mutual Fund Behavior?". *Journal of Finance*. 63: 2941–2969.

Phillips, B., K. Pukthuanthong, and P. R. Rau. 2018. "Size Does Not Matter: Diseconomies Of Scale In The Mutual Fund Industry Revisited". *Journal of Banking and Finance*. 88: 357–365.

Reuter, J. and E. Zitzewitz. 2015. "How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach". Boston College Working Paper.

Rousseeuw, P. J. and B. C. van Zomeren. 1990. "Unmasking Multivariate Outliers And Leverage Points". *Journal of the American Statistical Association*. 85: 633–639.

Yan, X. S. 2008. "Liquidity, Investment Style, And The Relation Between Fund Size And Fund Performance". *Journal of Financial and Quantitative Analysis*. 43: 741–768.

Zhu, M. 2018. "Informative Fund Size, Managerial Skill, And Investor Rationality". *Journal of Financial Economics*. 130(1): 114–134.

Panel A: CRSP and Morningstar Matched Funds Return Differences

	Published PST (2015) Data Appendix (1979–2011)		Our Replication (1979–2011)		Our Replication (1993–2011)		Our Main Sample (1993–2014)	
Difference (basis points)	Number of Observations	% of Observations	Number of Observations	% of Observations	Number of Observations	% of Observations	Number of Observations	% of Observations
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Do Not Differ	2,534,092	88.5%	2,291,849	91.10%	2,245,880	91.30%	2,912,570	92.01%
Differ = 1	82,507	2.9%	66,944	2.67%	66,099	2.69%	84,792	2.68%
2 ≤ Differ < 11	156,677	5.5%	118,836	4.72%	113,552	4.62%	130,385	4.11%
11 ≤ Differ ≤ 100	74,676	2.6%	33,475	1.33%	30,383	1.24%	33,170	1.02%
Differ > 100	14,001	0.5%	0	0%	0	0%	0	0%

Panel B: Descriptive Statistics of the Merged CRSP-Morningstar Data

	Obs.	Mean	St. Dev.	MIN.	1%	25%	Percentile 50%	75%	99%	MAX.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Published (1979–2011)										
Benchmark-adj. Gross Return	314,580	0.0005	0.0229	N.A.	–0.0627	–0.0095	–0.0009	0.0100	0.0680	N.A.
FUNDSIZE	319,454	1,564	5,779	N.A	16	84	265	921	24,371	N.A.
INDUSTRYSIZE	381,192	0.1334	0.0441	N.A.	0.0232	0.1227	0.1391	0.1714	0.1833	N.A.
Our Replication (1979–2011)										
Benchmark-adj. Gross Return	293,264	–0.0002	0.0226	–0.3169	–0.0625	–0.0098	–0.0006	0.0088	0.0680	0.4205
FUNDSIZE	293,264	1,668	5,979	15	18	98	300	1,005	25,442	180,759
INDUSTRYSIZE	293,264	0.1637	0.0330	0.0011	0.0530	0.1418	0.1735	0.1909	0.2003	0.2030
Our Replication (1993–2011)										
Benchmark-adj. Gross Return	285,863	–0.0002	0.0227	–0.3169	–0.0229	–0.0098	–0.0006	0.0088	0.0683	0.4205
FUNDSIZE	285,863	1,629	5,965	15	18	95	292	969	23,360	180,758
INDUSTRYSIZE	285,863	0.1667	0.0274	0.0740	0.0906	0.1433	0.1741	0.1911	0.2003	0.2030
Our Main Sample (1993–2014)										
Benchmark-adj. Gross Return	347,710	–0.0003	0.0213	–0.3169	–0.0596	–0.0093	–0.0006	0.0081	0.0640	0.4205
FUNDSIZE	347,710	2,576	9,437	15	26	145	456	1,550	40,471	290,677
INDUSTRYSIZE	347,710	0.1698	0.0258	0.0740	0.0921	0.1500	0.1824	0.1900	0.2003	0.2030

Table 1: Statistics of the Merged CRSP-Morningstar Data.

Description: This table reports the statistics of the matched CRSP-Morningstar dataset. Panel A details the matching statistics published in the PST online data appendix in columns 2 and 3 from 1979 to 2011 and our replication sample using the PST sample period in columns 4 and 5. Columns 6 and 7 report the matching statistics over the period March 1993 to December 2011 and columns 8 and 9 report the matching statistics for our sample from March 1993 to December 2014. Panel B reports descriptive statistics of the merged data. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio-level total net assets at the end of the previous month (inflated to 2011 or 2014 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks.

Interpretation: Our replication sample are nearly identical the original PST.

	Published			Replication		
	OLS	OLS FE	RD	OLS	OLS FE	RD
	(1)	(2)	(3)	(4)	(5)	(6)
FUNDSIZE	–0.015	–0.148	–0.425	–0.017	–0.121	–0.067
	(2.02)	(9.09)	(1.25)	(3.15)	(8.87)	(0.99)
INDUSTRYSIZE	–0.017	–0.030	–0.028	–0.011	–0.018	–0.021
	(1.90)	(3.27)	(2.14)	(1.71)	(2.74)	(2.40)
Constant	0.003			0.002		
	(2.09)			(1.44)		
Observations	275,847	275,847	270,556	277,346	277,346	275,651

Table 2: PST (2015) and Our Results for Same Period (March 1993–December 2011).

Description: This table compares the PST results with our replicated results. This table replicates Table 3, Panel A, columns 7, 8, and 9 in PST using the merged CRSP-Morningstar database we generate following the online Data Appendix for the published paper and report results using Morningstar benchmark-adjusted gross returns at the portfolio level from the merged CRSP-Morningstar databases as the dependent variable. The sample period for the published and replicated results is March 1993 to December 2011. FUNDSIZE is the portfolio level of total net assets at the end of the previous month inflated to December 2011 using the total market capitalization of stocks in the CRSP stock database. INDUSTRYSIZE is the proportion of the dollar value of all actively managed mutual funds to the total market capitalization of all stocks in the CRSP stock database. The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels. The *t*-statistics for the recursive demeaning (RD) regressions are also adjusted for clustering at the fund level.

Interpretation: Our regression results are nearly identical to the original PST.

	Mean (Median) Values			Mean (Median) Differences	
	Non-Outlier (1)	Vertical (2)	Horizontal (3)	Vertical – Non-Outlier (2) – (1)	Horizontal – Non-Outlier (3) – (1)
Benchmark-adj.	–0.0007	0.0026	0.0032	0.0033[a]	0.0039[a]
Gross Return	(–0.0007)	(0.0301)	(0.0300)	(0.0308)[a]	(0.0307)[a]
FUNDSIZE	1,517	247	2,545	–1,270[a]	1,028[a]
	(245)	(174)	(1,077)	(–71)[a]	(832)[a]
INDUSTRYSIZE	0.1744	0.1662	0.1554	–0.0088[a]	–0.0196[a]
	(0.1865)	(0.1582)	(0.1466)	(–0.0283)[a]	(0.0399)[a]
Observations	222,448	17,310	17,484		

Table 3: Means and Medians by Outlier Type (1993–2014).

Description: This table reports mean and median values and provides segmentations for each observation type as well as mean and median differences in the non-outlier and outlier values. The sample is from March 1993 to December 2014. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio-level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. The notation [a] denotes statistical significance at the 1% level.

Interpretation: Potential data errors vary significantly in returns and scale.

Panel A: OLS Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.013	–0.013	–0.014	–0.016	–0.020	–0.023	–0.024	0.012	0.092	0.228
	(2.50)	(2.28)	(2.24)	(2.12)	(2.04)	(1.87)	(1.62)	(0.59)	(3.53)	(6.40)
INDUSTRYSIZE	–0.014	–0.013	–0.012	–0.011	–0.007	–0.003	0.004	0.018	0.033	0.051
	(2.28)	(2.10)	(1.98)	(1.79)	(1.26)	(0.53)	(0.68)	(3.25)	(5.97)	(9.16)
Observations	334,275	334,191	334,107	333,939	333,439	332,603	330,935	325,935	317,695	301,478
Observations Removed	0	84	168	336	836	1,672	3,340	8,340	16,580	32,797
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.49%	4.96%	9.81%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.76%	24.89%
Horizontal Outliers (%)	N.A.	50.00%	50.60%	50.30%	50.36%	50.60%	52.06%	53.37%	39.75%	28.49%
Good Outliers (%)	N.A.	50.00%	49.40%	49.70	49.64%	49.40%	47.93%	46.63%	46.61%	46.61%
R^2	0.03%	0.03%	0.02%	0.02%	<0.01%	<0.01%	<0.01%	0.06%	0.22%	0.60%

Panel B: OLS Fixed Effects Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.120	–0.155	–0.194	–0.232	–0.251	–0.298	–0.408	–0.562	–0.630	–0.705
	(9.38)	(9.51)	(9.59)	(9.47)	(9.66)	(9.92)	(12.09)	(13.24)	(12.89)	(10.31)
INDUSTRYSIZE	–0.022	–0.020	–0.019	–0.018	–0.013	–0.009	–0.001	0.016	0.033	0.054
	(3.28)	(3.07)	(2.94)	(2.76)	(2.24)	(1.42)	(0.11)	(2.77)	(5.72)	(9.26)
Observations	334,275	334,191	334,107	333,939	333,439	332,603	330,935	325,935	317,695	301,478
Observations Removed	0	84	168	336	836	1,672	3,340	8,340	16,580	32,797
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.49%	4.96%	9.81%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.76%	24.89%
Horizontal Outliers (%)	N.A.	50.00%	50.60%	50.30%	50.36%	50.60%	52.06%	52.98%	39.75%	28.49%
Good Outliers (%)	N.A.	50.00%	49.40%	49.70	49.64%	49.40%	47.93%	46.46%	46.61%	46.61%
R^2	1.12%	1.21%	1.27%	1.31%	1.60%	1.80%	2.06%	2.65%	4.20%	6.41%

Panel C: Recursive Demeaning Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.073	–0.285	0.458	–0.304	–0.308	–0.319	–0.365	–0.521	–0.264	-1.190
	(0.84)	(0.67)	(0.36)	(0.82)	(1.00)	(1.26)	(1.40)	(0.34)	(0.17)	(0.77)
INDUSTRYSIZE	–0.021	–0.017	–0.014	–0.013	–0.006	0.003	0.017	0.046	0.076	0.132
	(2.44)	(1.70)	(0.89)	(1.34)	(0.68)	(0.30)	(1.98)	(3.02)	(4.29)	(7.33)
Observations	332,516	332,432	332,348	332,182	331,682	330,850	329,187	324,214	315,999	299,864
Observations Removed	0	84	168	334	834	1,666	3,329	8,302	16,517	32,652
Actual Removed (%)	0.00%	0.03%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	4.97%	9.82%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.76%	24.84%
Horizontal Outliers (%)	N.A.	50.00%	50.60%	50.00%	50.24%	50.48%	51.97%	53.26%	39.68%	28.45%
Good Outliers (%)	N.A.	50.00%	49.40%	50.00%	49.76%	49.52%	48.03%	46.74%	46.56%	46.71%
R^2	<0.01%	<0.01%	<0.01%	<0.01%	0.07%	<0.01%	<0.01%	0.29%	0.53%	0.50%

Table 4: Scale and Fund Performance When Percentages of Outliers are Removed (1993–2014).

Description: This table presents results obtained from regressing fund gross benchmark-adjusted returns on FUNDSIZE and INDUSTRYSIZE with increasing (from left to right) levels of multivariate outliers removed for OLS (Panel A), OLS with fund fixed effects (Panel B), and recursive demeaning (RD) estimators (Panel C). Following PST, the RD estimator recursively forward demeans all variables and instruments for FUNDSIZE using backward-demeaned FUNDSIZE. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio-level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. This table compares to Table 3, Panel A, columns 7, 8, and 9 in PST but with the merged CRSP-Morningstar databases for the period March 1993 to December 2014 rather than the March 1993 to December 2011 period in the original paper. The variable Actual Removed is the ratio of observations removed over total observations of the full sample. The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels. The *t*-statistics are also adjusted for clustering at the fund level in the RD regressions (Panel C).

Interpretation: Diseconomies of scale at the industry level findings are driven by small numbers of potential data errors.

Panel A: Most Extreme Observations

Fund Name	Morningstar Category	Morningstar Benchmark Index	No. Months Top 0.05%	Good Outlier	Vertical Outlier	Horizontal Outlier
American Funds Growth Fund of America	US Fund Large Growth	Russell 1000 Growth	84	Yes	No	No
Comstock Capital Value Fund	US Fund Bear Market	S&P 500	7	No	No	Yes
Federated Prudent Bear Fund	US Fund Bear Market	S&P 500	6	No	No	Yes
Matthews Korea Fund	US Fund Miscellaneous Region	MSCI ACWI Ex USA	4	No	No	Yes
Voya Russia Fund	US Fund Miscellaneous Region	MSCI ACWI Ex USA	4	No	No	Yes
Grizzly Short Fund	US Fund Bear Market	S&P 500	4	No	No	Yes
Nuveen Mid Cap Select Fund	US Fund Mid-Cap Blend	Russell Mid Cap	3	No	No	Yes
Embarcadero Absolute Return	US Fund Small Growth	Russell 2000 Growth	3	No	No	Yes
Scout Small Cap Fund	US Fund Small Growth	Russell 2000 Growth	2	No	No	Yes
Brown Advisory Opportunity Fund	US Fund Mid-Cap Growth	Russell Mid Cap Growth	2	No	No	Yes
Embarcadero Market Neutral	US Fund Mid-Cap Growth	Russell Mid Cap Growth	2	No	No	Yes
Comstock Strategy Fund	US Fund Bear Market	S&P 500	2	No	No	Yes
CGM Focus Fund	US Fund Large Blend	Russell 1000	2	No	No	Yes
Victory Munder Multi-Cap Fund	US Fund Large Growth	Russell 1000 Growth	2	No	No	Yes
Nuveen Small Cap Growth Opps	US Fund Small Growth	Russell 2000 Growth	1	No	No	Yes
Al Frank Fund	US Fund Large Value	Russell 1000 Value	1	No	No	Yes
Caldwell & Orkin Market Opportunity Fund	US Fund Long-Short Equity	S&P 500	1	No	No	Yes
Century Shares Trust Fund	US Fund Large Growth	Russell 1000 Growth	1	No	No	Yes
Old Westbury Mid Cap Equity Fund	US Fund Mid-Cap Growth	Russell Mid Cap Growth	1	No	No	Yes
TCW Relative Value Small Cap Fund	US Fund Small Blend	Russell 2000	1	No	No	Yes

Table 5: Identifying Data Errors.

Description: This panel lists the 20 funds that most frequently have vertical and horizontal distances (large returns and size in a multivariate sense) in the top 0.05% and/or have the largest vertical and horizontal distances. The top 0.05% of observations with the most extreme vertical and horizontal distances includes 168 observations from 54 funds. All funds are in US dollars.

Interpretation: All of the most extreme 0.05% observations are data errors.

Panel B: Scale Regressions after Cleaning the Data

	Stage 1: Remove Current vs. Historical Category Errors		Stage 2: Remove Current vs. Historical Style Errors		Stage 3: Remove Morningstar, Benchmarking, and Sampling Errors	
	OLS FE	RD	OLS FE	RD	OLS FE	RD
	(1)	(2)	(3)	(4)	(5)	(6)
FUNDSIZE	–0.119	–0.051	–0.120	–0.056	–0.260	–0.239
	(9.61)	(0.58)	(9.65)	(0.62)	(9.80)	(1.59)
INDUSTRYSIZE	–0.013	–0.007	–0.013	–0.006	–0.011	–0.002
	(2.12)	(0.80)	(2.05)	(0.69)	(1.83)	(0.27)
Data Errors	679	679	74	74	289	289
Cumulative	679	679	753	753	1,042	1,041
% of Sample	0.20%	0.20%	0.23%	0.23%	0.31%	0.31%
% of Top 0.5%	40.61%	40.76%	45.04%	45.20%	62.32%	62.48%
Observations	333,596	331,839	333,522	331,765	333,233	331,477
R^2	0.88%	<0.01%	<0.88%	<0.01%	<0.01%	<0.01%

Table 5: Identifying Data Errors.

Description: This panel reports results from regressing gross Morningstar benchmark-adjusted returns on the variables FUNDSIZE and INDUSTRYSIZE using the Fund Fixed Effects and PST recursive demeaning procedures. We clean the data in the top 0.5% (1,672 observations) and rerun the regressions in three stages: (i) dropping Morningstar Category errors that occur when the historical-observation period and the end of sample Morningstar Category differ (columns 1 and 2), (ii) dropping Morningstar Style errors that occur when the historical style assignment differs from the end of sample Morningstar Category Style (columns 3 and 4), and (iii) dropping Morningstar Category and Style, Benchmarking, and sampling errors (columns 5 and 6). The variables are from our merged CRSP-Morningstar mutual fund database that is constructed following PST. The sample period for the published and replicated results is January 2002 to December 2014. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio level of total net assets at the end of the previous month inflated to December 2014 using the total market capitalization of stocks in the CRSP stock database. INDUSTRYSIZE is the proportion of the dollar value of all actively managed mutual funds to the total market capitalization of all stocks in the CRSP stock database. The fixed effect model *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels. The *t*-statistics are also adjusted for clustering at the fund level in the recursive demeaning regressions.

Interpretation: Insignificant returns to industry scale after cleaning the data to remove data errors that should not have been in the sample.

Year	Number of Top 0.05% Most Extreme	Number of Top 0.50% Most Extreme
1993	0	10
1994	0	12
1995	0	13
1996	0	32
1997	0	56
1998	2	67
1999	11	156
2000	45	380
2001	12	196
2002	5	82
2003	0	42
2004	0	38
2005	5	50
2006	12	58
2007	12	66
2008	19	107
2009	14	100
2010	12	51
2011	12	44
2012	7	41
2013	0	36
2014	0	35

Table 6: Most Extreme Observations and Data Errors by Year.

Description: This panel provides the distribution of the top 0.05% and 0.50% extreme observations by year. Data generated following PST for the March 1993 to December 2014 period.

Interpretation: Extreme observations and data errors occur most often during periods of heighted market uncertainty.

	No Dropping OLS FE	Drop 0.50% Most Extreme Outliers OLS FE	No Dropping RD	Drop 0.05% Most Extreme Outliers RD
	(1)	(2)	(3)	(4)
FUNDSIZE	–0.167	–0.348	–0.051	0.361
	(11.57)	(11.54)	(0.64)	(0.23)
INDUSTRYSIZE	–0.012	–0.012	–0.013	–0.015
	(0.86)	(0.87)	(0.91)	(0.71)
Observations	244,736	243,990	240,031	239,910
R^2	0.44%	1.20%	<0.01%	<0.01%

Table 7: Scale and Fund Performance Post 2001 (2002–2014).

Description: This table reports results from regressing gross Morningstar benchmark-adjusted returns on the variables FUNDSIZE and INDUSTRYSIZE using the PST recursive demeaning procedure. The variables are from our merged CRSP-Morningstar mutual fund database that is constructed following PST. The sample period for the published and replicated results is January 2002 to December 2014. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio level of total net assets at the end of the previous month inflated to December 2014 using the total market capitalization of stocks in the CRSP stock database. INDUSTRYSIZE is the proportion of the dollar value of all actively managed mutual funds to the total market capitalization of all stocks in the CRSP stock database. The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels. The *t*-statistics are also adjusted for clustering at the fund level in the recursive demeaning regressions.

Interpretation: Returns are unrelated to scale post 2001 in the full and dropped samples.

Panel A: OLS Fixed Effects Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.110	–0.141	–0.175	–0.211	–0.223	–0.263	–0.378	–0.543	–0.613	–0.671
	(10.56)	(10.47)	(10.44)	(10.48)	(10.40)	(10.47)	(12.65)	(14.45)	(14.47)	(12.13)
INDUSTRYSIZE	–0.013	–0.013	–0.012	–0.010	–0.008	-.005	0.001	0.011	0.021	0.032
	(1.72)	(1.60)	(1.54)	(1.41)	(1.12)	(0.70)	(0.16)	(1.64)	(3.29)	(5.36)
Observations	334,271	334,187	334,103	333,935	333,435	332,599	330,928	325,933	317,690	301,483
Observations Removed	0	84	168	336	836	1,672	3,343	8,338	16,581	32,788
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.49%	4.96%	9.81%
R^2	2.00%	0.54%	0.53%	0.51%	0.61%	0.77%	1.06%	1.72%	2.69%	4.39%

Panel B: OLS Fixed Effects Regression Estimates After Removing Bear Market, Market Neutral, Specialized Sector (IPO and Internet), and International Funds

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.109	–0.140	–0.174	–0.209	–0.221	–0.260	–0.374	–0.536	–0.603	–0.661
	(10.54)	(10.40)	(10.34)	(10.38)	(10.29)	(10.36)	(12.55)	(14.37)	(14.56)	(11.98)
INDUSTRYSIZE	–0.013	–0.012	–0.012	–0.011	–0.008	–0.005	0.001	0.011	0.021	0.032
	(1.62)	(1.56)	(1.51)	(1.40)	(1.12)	(0.70)	(0.14)	(1.63)	(3.26)	(5.30)
Observations	329,152	329,084	329,008	328,853	328,387	327,590	325,973	321,112	313,020	297,059
Observations Removed	0	68	144	299	765	1,562	3,179	8,040	16,132	32,093
Actual Removed (%)	0.00%	0.021%	0.04%	0.09%	0.23%	0.48%	0.97%	2.44%	4.90%	9.75%
R^2	2.08%	0.56%	0.53%	0.53%	0.60%	0.73%	0.94%	1.46%	2.26%	3.75%

Panel C: Recursive Demeaning Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.095	–0.543	–0.525	–0.334	–0.306	–0.297	–0.451	-3.227	2.079	–0.262
	(1.36)	(0.33)	(0.34)	(0.77)	(0.92)	(1.26)	(1.82)	(0.062)	(0.71)	(0.19)
INDUSTRYSIZE	–0.013	–0.008	–0.007	–0.007	–0.003	0.002	0.014	0.058	0.033	0.091
	(1.44)	(0.40)	(0.39)	(0.70)	(0.35)	(0.26)	(1.62)	(1.19)	(1.14)	(5.84)
Observations	332,511	332,428	332,345	332,177	331,678	330,845	329,182	324,208	316,003	299,877
Observations Removed	0	83	166	334	833	1,666	3,329	8,303	16,508	32,634
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	4.96%	9.81%
R^2	<0.01%	<0.01%	<0.01%	0.02%	<0.01%	<0.01%	<0.01%	0.06%	0.22%	1.69%

Table 8: Scale and Fama-French Adjusted Fund Performance.

Description: This table presents results obtained from regressing fund gross Fama-French adjusted returns on FUNDSIZE and INDUSTRYSIZE with increasing (from left to right) levels of multivariate outliers removed for OLS with fund fixed effects (Panel A), OLS with fund fixed effects and misclassified funds removed (Panel B), and recursive demeaning (RD) estimators (Panel C). Following PST, the RD estimator recursively forward demeans all variables and instruments for FUNDSIZE using backward-demeaned FUNDSIZE. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio-level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. This table compares to Table 3, Panel A, columns 7, 8, and 9 in PST but with the merged CRSP-Morningstar databases for the period March 1993 to December 2014 rather than the March 1993 to December 2011 period in the original paper. The variable Actual Removed is the ratio of observations removed over total observations of the full sample. The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels. The *t*-statistics are also adjusted for clustering at the fund level in the RD regressions (Panel C).

Interpretation: The no relation between industry scale and performance findings are robust to Fama-French alphas where benchmarking errors are not relevant.

	Percentage of Sample Dropped									
	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Dollar AUM										
FUNDSIZE	–0.0824	–0.0893	–0.1018	–0.1346	–0.1803	–0.1925	–0.3423	–0.2921	0.2946	4.9593
	(1.45)	(1.09)	(1.02)	(1.21)	(1.13)	(1.36)	(1.40)	(1.76)	(0.80)	(3.23)
INDUSTRYSIZE	–0.0153	–0.0128	–0.0110	–0.0077	–0.0004	0.0084	0.0242	0.0520	0.0788	0.0679
	(1.51)	(1.27)	(1.09)	(0.77)	(0.04)	(0.86)	(2.34)	(5.61)	(8.00)	(2.40)
Observations	315,167	315,087	315,009	314,851	314,379	313,591	312,017	307,305	299,539	284,246
Panel B: Ln Dollar AUM										
*ln*FUNDSIZE	–0.0018	–0.0018	–0.0018	–0.0018	–0.0017	–0.0015	–0.0013	–0.0007	–0.0002	0.0000
	(6.44)	(6.35)	(6.41)	(6.49)	(6.49)	(5.95)	(5.19)	(2.82)	(0.98)	(0.14)
INDUSTRYSIZE	–0.0109	–0.0096	–0.0089	–0.0077	–0.0044	0.0037	0.0145	0.0474	0.0852	0.1463
	(1.09)	(0.99)	(0.91)	(0.80)	(0.48)	(0.41)	(1.61)	(5.16)	(8.68)	(13.06)
Observations	313,581	313,522	313,466	313,347	312,995	312,236	310,720	306,122	298,558	283,762

Table 9: Scale and Fund Performance Using Zhu's (2018) Recursive Demeaning Estimator.

Description: The table presents results from regressing fund gross benchmark-adjusted returns on FUNDSIZE and INDUSTRYSIZE with increasing levels of multivariate outliers removed for FUNDSIZE measure in dollar AUM terms (Panel A) and in the natural logarithm of dollar AUM (Panel B) using Zhu (2018) recursive demeaning (RD) estimators. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database), *ln*FUNDSIZE is the natural logarithm of FUNDSIZE, and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. This table compares to Table 3, Panels A and B, column 4 in Zhu (2018) but with INDUSTRYSIZE as an added independent variable and with the merged CRSP-Morningstar databases for the period March 1993 to December 2014 instead of the Morningstar-only database for the period January 1995 to December 2014 as in Zhu (2018). The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector, month, and fund levels.

Interpretation: There is no relation between industry scale and benchmark adjusted performance. Fund scale results are sensitive to functional form (AUM or l*n*AUM).

Size/Return Quintiles	Typical Fund Observations Lowest 1	2	3	4	Highest 5	Measure	Extreme Observations Lowest 1	2	3	4	Highest 5
Smallest 1	–0.027	–0.008	–0.001	0.006	0.022	Mean	–0.128	–0.008	–0.0003	0.008	0.076
	(–0.022)	(–0.008)	(–0.001)	(0.006)	(0.020)	Median	(–0.128)	(–0.008)	(–0.0001)	(0.008)	(0.065)
	13,407	12,017	11,252	11,863	12,124	Obs.	174	25	20	19	1,816
2	–0.026	–0.008	–0.001	0.006	0.022	Mean	–0.056	–0.008	–0.001	0.006	0.067
	(–0.021)	(–0.007)	(–0.001)	(0.006)	(0.020)	Median	(–0.027)	(–0.008)	(–0.0004)	(0.006)	(0.063)
	12,602	12,370	12,020	12,134	11,401	Obs.	256	151	148	157	1,477
3	–0.026	–0.008	–0.001	0.006	0.022	Mean	–0.049	–0.008	–0.001	0.007	0.065
	(–0.021)	(–0.007)	(–0.001)	(0.006)	(0.019)	Median	(–0.024)	(–0.008)	(–0.001)	(0.007)	(0.060)
	12,253	12,171	12,365	12,317	11,153	Obs.	310	231	208	262	1,446
4	–0.025	–0.008	–0.001	0.006	0.022	Mean	–0.053	–0.008	–0.001	0.007	0.061
	(–0.020)	(–0.007)	(–0.001)	(0.006)	(0.019)	Median	(–0.028)	(–0.007)	(–0.001)	(0.007)	(0.058)
	11,438	12,385	13,098	12,830	10,521	Obs.	398	238	213	263	1,332
Largest 5	–0.025	–0.008	–0.001	0.006	0.021	Mean	–0.049	–0.007	–0.001	0.007	0.057
	(–0.021)	(–0.007)	(–0.001)	(0.006)	(0.019)	Median	(–0.027)	(–0.007)	(–0.001)	(0.006)	(0.055)
	11,457	12,836	13,102	12,483	9,986	Obs.	422	292	290	388	1,460
5-1	0.002^{a}	0.0001^{a}	0.000	–0.0002^{a}	–0.001^{a}	Mean	0.079^{a}	0.0004	–0.0004	–0.001	–0.019^{a}
	(0.001)^{a}	(0.0002)^{a}	(0.000)	(–0.0003)^{a}	(–0.001)^{a}	Median	(0.101)^{a}	(0.001)^{b}	(–0.001)	(–0.002)	(–0.010)^{a}
	-1,950	819	1,850	620	-2,138	Obs.	248	267	270	369	-356

Table 10: Two-Way Sorts of Fund Size and Return Quintiles.

Description: The table presents mean (median) fund gross benchmark-adjusted returns and the number of fund-month observations in two-way sort of gross benchmark-adjusted returns and one-month-lagged fund size for the typical and influential observations. Extreme observations are the minimum number of most extreme outliers whose removal causes the Zhu (2018) recursive demeaning (RD) coefficient estimates on FUNDSIZE to become insignificant when regressing gross benchmark-adjusted returns on *ln*FUNDSIZE and INDUSTRYSIZE. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database), *ln*FUNDSIZE is the natural logarithm of FUNDSIZE, and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. These data are from the merged CRSP-Morningstar database for the period March 1993 to December 2014. The notations ^{a} and ^{b} denote statistical significance at the 1% and 5% levels.

Interpretation: Relations between returns and scale occur primarily in suspected data errors (extreme observations) and vary by portfolio size.

Panel A: Regressions by Quintiles

	Regressions by Quintiles						
	Smallest Quintile (1)	2	3	4	Largest Quintile (5)	Quintiles 1&2	Quintiles 3&4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
*ln*FUNDSIZE	–0.0022	–0.0066	0.0070	0.0038	0.0012	–0.0027	0.0028
	(4.97)	(3.70)	(2.38)	(2.78)	(1.29)	(6.30)	(2.62)
INDUSTRYSIZE	–0.0113	0.0450	–0.0047	0.0075	0.0171	0.0039	–0.0105
	(0.44)	(2.40)	(2.07)	(0.66)	(1.39)	(0.27)	(1.06)
Observations	62,717	62,716	62,716	62,716	62,716	125,433	125,432

Panel B: Sensitivity to Outliers by Quintile

	Percentage of Sample Dropped									
	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Quintiles 1and 2										
*ln*FUNDSIZE	–0.0027	–0.0027	–0.0029	–0.0026	–0.0024	–0.0021	–0.0017	–0.0008	0.0001	0.0013
	(6.30)	(6.32)	(6.27)	(6.25)	(6.09)	(5.58)	(4.77)	(2.82)	(0.45)	(3.51)
INDUSTRYSIZE	0.0039	0.0049	0.0062	0.0072	0.0091	0.0167	0.0266	0.0602	0.0950	0.1403
	(0.27)	(0.34)	(0.44)	(0.52)	(0.67)	(1.28)	(2.13)	(4.90)	(7.20)	(9.43)
Observations	125,433	125,404	125,373	125,310	125,130	124,865	124,307	122,713	120,127	115,080
Observations Dropped	0	29	60	123	303	568	1,126	2,720	5,306	10,353
Actual % Obs. Dropped	0	0.023%	0.05%	0.10%	0.24%	0.45%	0.90%	2.17%	4.23%	8.25%
Quintiles 3 and 4										
*ln*FUNDSIZE	0.0028	0.0028	0.0028	0.0028	0.0028	0.0026	0.0021	0.0010	–0.0005	–0.0040
	(2.62)	(2.68)	(2.67)	(2.68)	(2.67)	(2.48)	(2.05)	(1.07)	(0.51)	(3.62)
INDUSTRYSIZE	–0.0105	–0.0103	–0.0092	–0.0080	–0.0063	0.0008	0.0103	0.0421	0.0779	0.1360
	(1.06)	(1.05)	(0.95)	(0.82)	(0.66)	(0.09)	(1.12)	(4.55)	(7.86)	(10.76)
Observations	125,432	125,409	125,393	125,355	125,247	124,933	124,350	122,442	119,296	112,953
Observations Dropped	0	23	39	77	185	499	1,082	2,990	6,136	12,479
Actual % Obs. Dropped	0	0.018%	0.03%	0.06%	0.15%	0.40%	0.86%	2.38%	4.89%	9.95%

Quintile 5										
*ln*FUNDSIZE	0.0012	0.0011	0.0011	0.0010	0.0009	0.0007	0.0002	–0.0005	–0.0020	–0.0045
	(1.29)	(1.23)	(1.24)	(1.17)	(1.01)	(0.81)	(0.28)	(0.59)	(2.45)	(5.17)
INDUSTRYSIZE	0.0171	0.0182	0.0177	0.0181	0.0204	0.0259	0.0302	0.0472	0.0565	0.0667
	(1.39)	(1.23)	(1.44)	(1.48)	(1.71)	(2.15)	(2.48)	(3.77)	(4.19)	(4.04)
Observations	62,716	62,709	62,700	62,682	62,618	62,438	62,063	60,967	59,135	55,729
Observations Dropped	0	7	16	34	98	278	653	1,749	3,581	6,987
Actual % Obs. Dropped	0	0.011%	0.03%	0.05%	0.16%	0.44%	1.04%	2.79%	5.71%	11.14%

Table 11: Scale and Fund Performance by Portfolio Size Quintiles.

Description: The table presents quintile-level results from regressing fund gross benchmark-adjusted returns on *ln*FUNDSIZE and INDUSTRYSIZE with increasing levels of multivariate outliers removed. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio-level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database), *ln*FUNDSIZE is the natural logarithm of FUNDSIZE, and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. This table compares to Table 4 in Zhu (2018) but with INDUSTRYSIZE as an added independent variable and with the merged CRSP-Morningstar databases for the period March 1993 to December 2014 period instead of the Morningstar-only database for the period January 1995 to December 2014 as in Zhu (2018). The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector, month, and fund levels.

Interpretation: Economies of scale at the fund and industry levels are fragile and inconsistent across portfolio sizes.

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1%	2.5%	5%	10%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.029	–0.002	0.005	0.011	0.059	0.127	0.126	0.292	0.511	0.885
	(0.59)	(0.03)	(0.11)	(0.21)	(1.04)	(2.22)	(2.49)	(4.61)	(7.01)	(10.11)
INDUSTRYSIZE	0.010	0.016	0.019	0.020	0.030	0.037	0.041	0.053	0.058	0.063
	(0.44)	(0.73)	(0.91)	(0.96)	(1.43)	(1.76)	(1.91)	(2.50)	(2.78)	(2.98)
Observations	25,256	25,248	25,242	25,230	25,192	25,128	25,002	24,635	24,012	22,814
Observations Removed	0	8	14	26	64	128	254	621	1,244	2,442
Actual Removed (%)	0.00%	0.03%	0.06%	0.10%	0.25%	0.51%	1.01%	2.46%	4.93%	9.67%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.92%	24.04%
Horizontal Outliers (%)	N.A.	100%	81.25%	72.00%	50.29%	75.37%	74.13%	65.47%	52.79%	33.82%
Good Outliers (%)	N.A.	0.00%	18.75%	28.00%	49.71%	24.63%	25.87%	34.53%	37.30%	41.45%
R^2	<0.01%	0.01%	0.01%	0.01%	0.03%	0.07%	0.07%	0.15%	0.30%	0.54%

Table 12: Scale and Passively-Managed Fund Performance (1993–2014).

Description: The table presents results from regressing index fund gross benchmark-adjusted returns on FUNDSIZE and INDUSTRYSIZE with increasing levels of multivariate outliers removed. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio level total net assets at the end of the previous month (inflated to 2014 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all passively managed mutual funds divided by the market capitalization of all CRSP stocks. This table compares to Table 3, Panel A, columns 7, 8, and 9 in PST but with the merged CRSP-Morningstar databases of index mutual funds for the period March 1993 to December 2014. The variable Actual Removed is the ratio of observations removed over total observations of the full sample. The t-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels.

Interpretation: There are positive economies of scale in passively managed monies.

		Economies of Scale						Fund Size
Research Paper (Listed by Year)	Pub. Status	Fund Size	Industry Size	Family Size	Methods	Data Source	Years Covered	Table Source
Chen, Hong, Huang, and Kubik (2004)	AER	Negative	NA	Positive	FM	CRSP	1963-99	T3
Edelen, Evans, and Kadlec (2007)	WP	Constant	NA	NA	FM	Morningstar/CRSP	1995-05	T5
Pollet and Wilson (2008)	JF	Negative	NA	Positive	FM	CRSP-Thomson	1975-00	T7
Yan (2008)	JFQA	Negative	NA	Positive	FM	CRSP-Thomson	1993-02	T5,T8
Bhojraj, Cho, and Yehuda (2012)*	JAR	Negative	NA	Constant	FM	CRSP	1992-99	T2
Elton, Gruber, and Blake (2012)	RAPS	Constant	NA	Constant	Multi-Index	CRSP	1999-09	T8
Ferreira, Keshwani, Miguel, and Ramos (2013)**	RoF	Positive	NA	Positive	FM	Lipper	1997-07	T5
Busse, Jiang, and Tang (2021)	WP	Negative	NA	Positive	FM	CRSP-Thomson	1980-12	T8
Pástor, Stambaugh, and Taylor (2015)	JFE	Constant	Negative	Constant	RD	CRSP/Morningstar	1993-11	T3,T9
Golez and Shive (2015)	WP	Negative	NA	NA	FM	CRSP	2000-15	T2,T3
Reuter and Zitzewitz (2015)	WP	Constant	NA	Constant	RDD	Morningstar Principia	1996-09	T6,A1
Harvey and Liu (2017)	WP	Negative	Negative	NA	EM	CRSP	1991-11	T5
Phillips, Pukthuanthong, and Rau (2018)*	JBF	Constant	NA	Negative	IV	CRSP	1992-10	T4
Adams, Hayunga, and Mansi (2018)	CFR	Constant	NA	Constant	FM	CRSP	1963-14	T8
Hong and Jiang (2018)	CFR	Negative	NA	Constant	FM	CRSP-Thomson	1981-16	T1
Zhu (2018)***	JFE	Negative	Negative	NA	RD	Morningstar	1995-14	T3-A2

Table 13: Summary of the Literature.

Description: This table summarizes the literature on returns to scale in active management. Publications are: American Economic Review (AER), Journal of Finance (JF), Journal of Financial Economics (JFE), Journal of Banking and Finance (JBF), Critical Finance Review (CFR), Journal of Accounting Research (JAR), Journal of Financial and Quantitative Analysis (JFQA), Review of Finance (RoF), Review of Asset Pricing Studies (RAPS), and working papers (WP). NA denotes not applicable. The methods used are Fama–MacBeth (FM), regression discontinuity design (RDD), recursive demeaning (RD), and expectation maximization (EM). The term NA denote not applicable (non-examined) by the authors.
* Bhojraj et al. (2012) find positive economies of scale but only before the year 2000 (i.e., before Regulation Full Disclosure).
** Ferreira et al. (2013) find positive economies of scale for non-U.S. funds and diseconomies of scale for U.S. funds.
*** Although Zhu (2018) documents evidence of diseconomies of scale at the industry level, her results were unreported.

Interpretation: The evidence on diseconomies of scale in the literature is mixed.

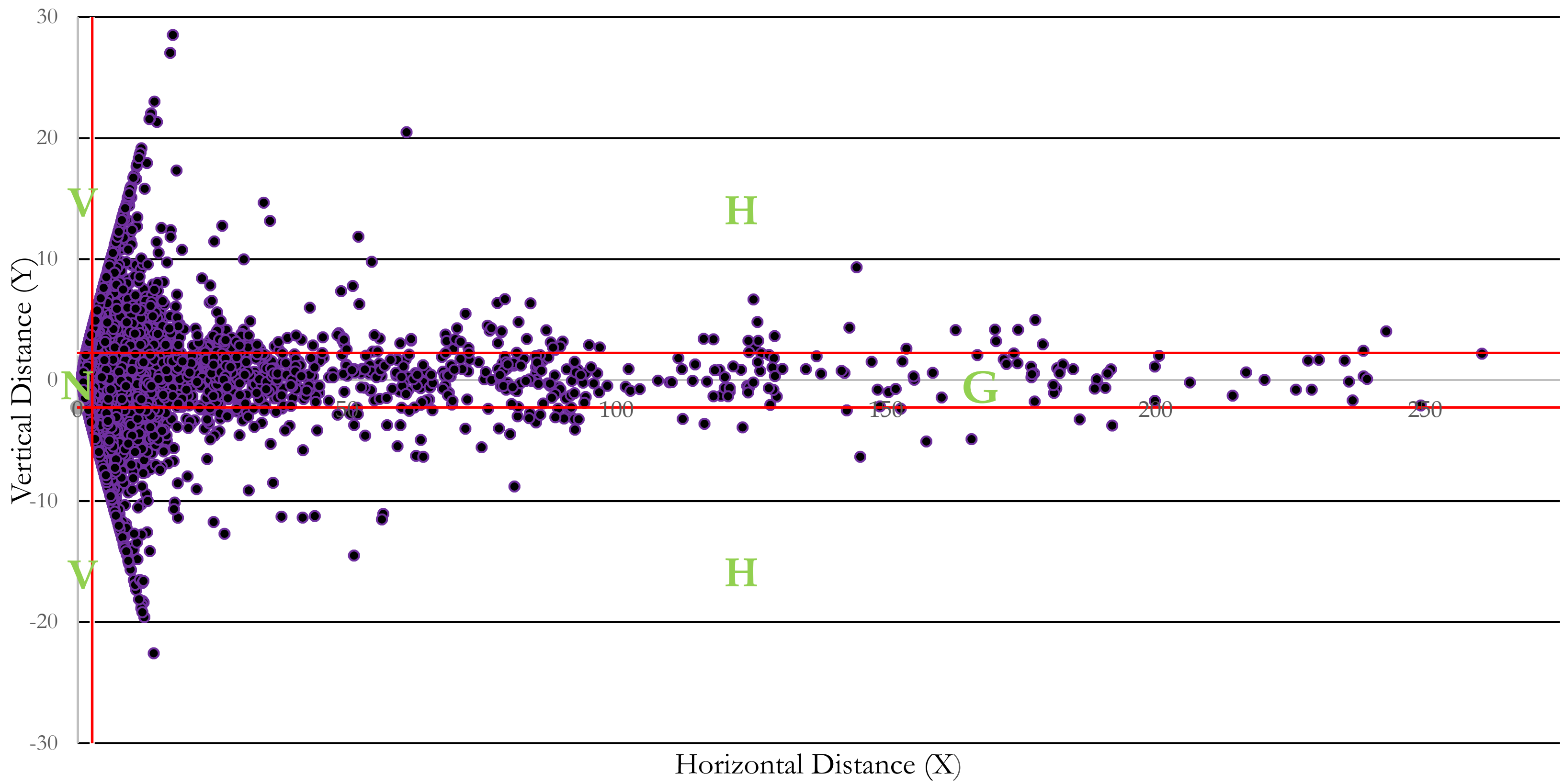

Figure 1. Vertical and Horizontal Distances.

Description: Outlier detection plot for the year 2000. Vertical distances (labeled "V") 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 +/- 2.25 are vertical outliers and are values from the standard normal distribution that separate the +/- 1.25% most remote regions from the central mass of observations. Observations with horizontal distances (labeled "H") to the right of the vertical boundary are horizontal outliers and located at $\chi^2_{p,0.975}$, where p is the number of parameters in the model and outside the horizontal bands. Good leverage points (labeled "G") are extreme in both dependent and independent variables space but fall on or near the regression line. The non-outlier funds, which constitute most of the observations, are located in region N.

Interpretation. This figure shows the sample contains unusual mutual funds that have the potential to bias coefficient estimates.

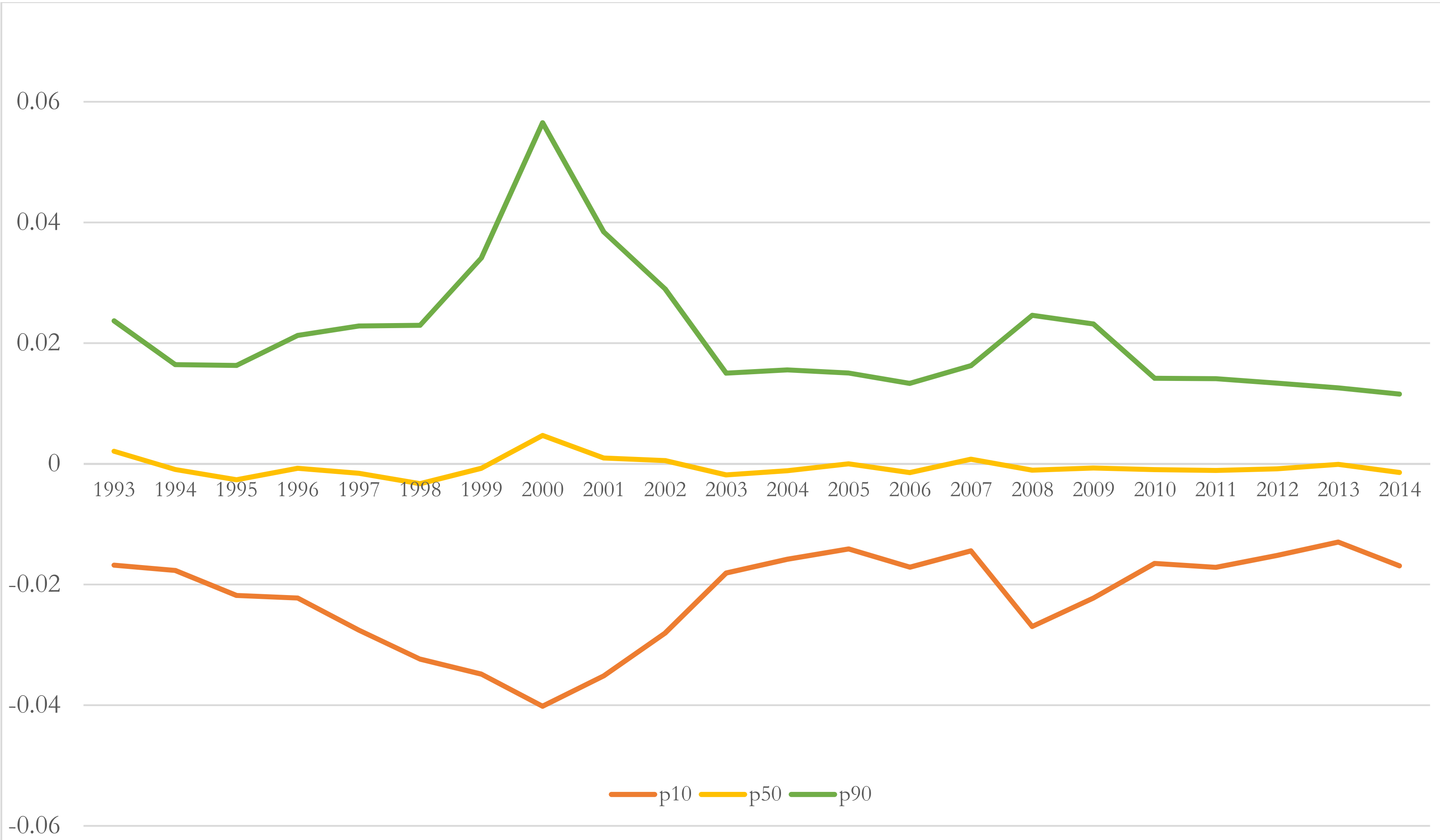

Figure 2. Percentiles of Monthly Gross Benchmark-adjusted Returns by Year: 1993 to 2014.

Description: Depiction of the 10th, 50th, and 90th percentiles of monthly gross benchmark-adjusted returns for the March 1993 to December 2014 period. The data are from our merged CRSP-Morningstar mutual fund database that is constructed following PST. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio.

Interpretation. Data errors, especially benchmarking errors, are more extreme in periods of increased market volatility.

	Mean (Median) Values			Mean (Median) Differences	
	Non-Outlier	Vertical	Horizontal	Vertical – Non-Outlier	Horizontal – Non-Outlier
	(1)	(2)	(3)	(2) – (1)	(3) – (1)
Benchmark-adj.	–0.0007	0.0013	0.0069	0.0020[a]	0.0075[a]
Gross Return	(–0.0007)	(0.0320)	(0.0341)	(0.0327)[a]	(0.0348)[a]
FUND SIZE	1,482	243	2,649	-1,240[a]	1,166[a]
	(246)	(175)	(1,144)	(-71)[a]	(898)[a]
INDUSTRY SIZE	0.1714	0.1622	0.1548	–0.0092[a]	–0.0166[a]
	(0.1786)	(0.1537)	(0.1466)	(–0.0249)[a]	(0.0320)[a]
Observations	182,767	16,151	16,925		

Appendix A: Means and Medians by Outlier Type (1993 to 2011).

Description: The table reports mean and median values and provides segmentations for each observation type as well as mean and median differences in the non-outlier and outlier values. The sample is from March 1993 to December 2011. Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio level total net assets at the end of the previous month (inflated to 2011 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. The notation [a] denotes statistical significance at the 1% level.

Interpretation: Potential data errors vary significantly in returns and scale for the PST sample period.

Panel A: OLS Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.017	–0.019	–0.020	–0.022	–0.028	–0.034	–0.042	–0.024	0.054	0.158
	(3.15)	(3.03)	(2.70)	(2.72)	(2.70)	(2.56)	(-2.62)	(1.02)	(1.86)	(3.78)
INDUSTRYSIZE	–0.011	–0.010	–0.009	–0.008	–0.005	–0.002	0.004	0.016	0.028	0.044
	(1.71)	(1.57)	(1.46)	(1.31)	(0.88)	(0.30)	(0.65)	(2.74)	(5.04)	(7.70)
Observations	277,346	277,276	277,206	277,068	276,652	275,958	274,574	270,429	263,582	250,125
Observations Removed	0	70	140	278	694	1,388	2,772	6,917	13,764	27,221
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.49%	4.96%	9.81%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.74%	24.40%
Horizontal Outliers (%)	N.A.	50.00%	50.00%	50.36%	50.29%	50.65%	52.20%	52.93%	39.60%	28.83%
Good Outliers (%)	N.A.	50.00%	50.00%	49.64	49.71%	49.35%	47.80%	47.07%	46.66%	46.76%
R^2	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	0.05%	0.16%	0.43%

Panel B: OLS Fixed Effects Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.120	–0.156	–0.210	–0.250	–0.268	–0.331	–0.441	–0.625	–0.742	–0.867
	(9.38)	(8.97)	(8.58)	(8.99)	(9.04)	(9.45)	(11.47)	(12.50)	(12.64)	(10.35)
INDUSTRYSIZE	–0.022	–0.017	–0.016	–0.015	–0.012	–0.007	–0.002	0.014	0.029	0.048
	(3.28)	(2.56)	(2.40)	(2.23)	(1.81)	(1.15)	(0.05)	(2.33)	(4.90)	(7.97)
Observations	277,346	277,276	277,206	277,068	276,652	275,958	274,574	270,429	263,582	250,125
Observations Removed	0	70	140	278	694	1,388	2,772	6,917	13,764	27,221
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.49%	4.96%	9.81%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.74%	24.40%
Horizontal Outliers (%)	N.A.	50.00%	50.00%	50.36%	50.29%	50.65%	52.20%	52.93%	39.60%	28.83%
Good Outliers (%)	N.A.	50.00%	50.00%	49.64	49.71%	49.35%	47.80%	47.07%	46.66%	46.76%
R^2	1.12%	1.22%	1.27%	1.33%	1.54%	1.74%	2.06%	2.83%	3.04%	6.03%

Panel C: Recursive Demeaning (RD) Regression Estimates

	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	5.00%	10.00%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
FUNDSIZE	–0.067	–0.270	0.420	–0.256	–0.264	–0.280	–0.298	–0.255	-2.326	-1.474
	(0.99)	(1.39)	(0.87)	(1.30)	(1.40)	(1.34)	(1.49)	(0.34)	(0.33)	(1.06)
INDUSTRYSIZE	–0.021	–0.016	–0.015	–0.014	–0.008	–0.001	0.012	0.044	0.088	0.124
	(2.40)	(1.78)	(1.44)	(1.55)	(0.91)	(0.01)	(1.54)	(5.06)	(1.23)	(7.49)
Observations	275,651	275,234	275,511	275,374	274,959	274,268	272,889	268,079	261,944	248,544
Observations Removed	0	70	140	278	694	1,388	2,772	6,917	13,764	27,221
Actual Removed (%)	0.00%	0.025%	0.05%	0.10%	0.25%	0.50%	1.00%	2.50%	4.99%	9.88%
Vertical Outliers (%)	N.A.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.73%	23.36%
Horizontal Outliers (%)	N.A.	50.00%	50.00%	50.00%	50.18%	50.54%	52.10%	52.83%	39.53%	28.80%
Good Outliers (%)	N.A.	50.00%	50.00%	50.00%	49.82%	49.46%	47.90%	47.16%	46.74%	46.84%
R^2	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%	1.32%

Appendix B: Scale and Performance When Percentages of Outliers are Removed (1993 to 2011).

Description: The table presents results from regressing fund gross benchmark-adjusted returns on FUNDSIZE and INDUSTRYSIZE with increasing levels of multivariate outliers removed for OLS (Panel A), OLS with fund fixed effects (Panel B), and recursive demeaning (RD) estimators (Panel C). Benchmark-adjusted gross returns are the monthly mutual fund net returns less the return on the Morningstar benchmark index portfolio plus 1/12 of the annual expense ratio. FUNDSIZE is the portfolio-level total net assets at the end of the previous month (inflated to 2011 values using the total market capitalization of stocks in the CRSP stock database) and INDUSTRYSIZE is the total net assets of all actively managed mutual funds divided by the market capitalization of all CRSP stocks. This table compares to Table 3, Panel A, columns 7, 8, and 9 in PST but with the merged CRSP-Morningstar databases for the period March 1993 to December 2011 as in the original paper. The variable Actual Removed is the ratio of observations removed over total observations of the full sample. The *t*-statistics in parentheses are adjusted for clustering at the Morningstar sector and month levels. The *t*-statistics are also adjusted for clustering at the fund level in the RD regressions (Panel C).

Interpretation: Diseconomies of scale at the industry level findings are driven by small numbers of potential data errors in the PST sample period.