

A Response to “Diseconomies of Scale in the Actively-Managed Mutual Fund Industry: What Do the Outliers in the Data Tell Us?” by Adams, Hayunga, and Mansi

Harrison Hong*

Wenxi Jiang†

Adams, Hayunda, and Mansi (2018, AHM hereafter) argue that the negative relationship between mutual fund size and performance, documented by Chen, Hong, Huang, and Kubik (2004, CHHK hereafter), is not robust to the removal of outliers in fund returns. AHM use an S-estimation method (Croux, Rousseeuw, and Hössjer 1994) to account for outliers and find that the point estimate for the effect of fund size is largely reduced as a result.¹ AHM further find that these return outliers are due to non-equity funds being misclassified as domestic equity funds in the 2004 Version of the CRSP Mutual Fund Database. These misclassified funds can be outliers and bias the point estimation since the performance adjustment method in CHHK was for domestic equity funds.

We agree with AHM that this problematic aspect of the CRSP Mutual Fund Database should be handled carefully. We disagree that we should simply implement robust regression methods to downweight outliers in fund returns. Extreme observations in fund returns contain valuable information about performance and its relationship to fund size. Since mutual fund returns are audited, there are no issues with the fund returns per se. It is simply an issue of cleaning up the classification of fund styles in the databases. Downweighting extreme positive or negative returns without a theory can also severely bias the estimate of average fund performance and fund size. Consider a world where small funds have positively skewed returns and large funds have negatively skewed returns, which can be consistent with diseconomies of scale. Downweighting performance at the tails would naturally also downward bias the estimate of the relationship between size and performance.

The ideal way to deal with this database style classification issue is to clean up the data as much as we can before even considering applying outlier filters. Since the 2004 publication

[‡] Columbia University and NBER; Email: hh2679@columbia.edu.

[†] CUHK Business School, The Chinese University of Hong Kong; Email: wenxijiang@baf.cuhk.edu.hk.

¹ There are many different robust regression models that downweight the influence of outliers on coefficients. The S-estimation method is one type. Results do depend on the type of robust regression method used.

of CHHK, the literature on mutual fund performance has come to recognize this issue and propose different ways to address it. These recent papers still find the fund size and performance results of CHHK. We show in this response that once the data is appropriately cleaned, the fund size–return relationship remains even if one applies outlier filters.

One way to clean the data, as in Pástor, Stambaugh, and Taylor (2015, PST hereafter), is to use Morningstar style information in conjunction with CRSP fund style information. Specifically, they identify actively-managed domestic equity-only mutual funds from the United States from 1979 to 2011 by combining CRSP with the fund style and benchmark information in Morningstar. They exclude index funds and very small funds, and require the information of fund size and return in CRSP and Morningstar to not to be very different. With this refined sample and OLS specifications, PST finds that the point estimate of fund size on performance is significant (page 34 and Table III, Panel B, column (1)). One difference of PST compared to CHHK is that PST uses dollar amount of fund size in regressions.

Another method, as in Harvey and Liu (2016), is to use fund holdings information and require funds to have 80% of their holdings in stocks. They use active, domestic equity funds covering the 1991-2011 period, consistent with PST. They start in 1991 as many funds do not have monthly updates on their TNA's before 1991. They only keep funds that have a TNA above \$10 million and have more than 80% of their holdings in stocks and require that a fund has at least 18 non-missing monthly observations. They use log transformation of TNA in the regression; the regression is specified as a structural estimation that features fund fixed effects. They find the negative relationship between fund size and performance remains both economically and statistically significant. (see Tables 4 and 5 of their paper)

We now show that, by using a refined list of domestic equity funds, the negative relationship between fund size and performance remain significant even using the S-estimation in the extended sample period. We identify equity funds by combining fund portfolio information. We download mutual funds' quarterly portfolio of domestic stocks from Thomson Reuters and merge it with CRSP via MFLINKS. To only keep the funds whose major investment are in U.S. equity, we require the market value of a fund's domestic stock portfolio to be higher than 75% of the fund's asset under management. This restriction is to address the potential misclassification in funds' self-reported style in CRSP. We also require the ratio not to

exceed 1.2, in order to avoid possible matching errors. To make the procedure transparent, we make as fewer changes as possible and keep the other screening conditions and specification the same as CHHK. The fund portfolio data is available from 1980, and we extend the sample period to 2016.

Table 1 presents the regression results. The left-hand variable is gross fund return. Columns (1)–(4) show the result of regressions as specified in CHHK. The coefficient of LOGTNA is significantly negative: the point estimate ranges from -0.027 to -0.035 with t -stats above 2.4. Note that this magnitude is larger to that shown in CHHK, i.e., -0.020 to -0.028. In columns (5)–(8), we use the S -estimation proposed by AHM and run the regressions as specified in Table 3 in AHM. The negative relationship remains in the refined equity fund sample: the coefficients are still sizeable, ranging from -0.016 to -0.027. The results using net fund returns are similar. Of course, the point estimates when using the S -estimator are smaller since we are imposing outlier controls and fund return outliers contain valuable information.

In sum, the negative relationship between fund returns and fund size of CHHK remains when one cleans up the CRSP Mutual Fund Database for fund style misclassifications. There is no need for outlier robust regressions per se. A careful cleaning of the data yields a more accurate judgment of this important relationship.

References

Chen, Joseph, et al. "Does fund size erode mutual fund performance? The role of liquidity and organization." *American Economic Review* 94.5 (2004): 1276-1302.

Croux, Christophe, Peter J. Rousseeuw, and Ola Hössjer. "Generalized S-estimators." *Journal of the American Statistical Association* 89.428 (1994): 1271-1281.

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor. "Scale and skill in active management." *Journal of Financial Economics* 116.1 (2015): 23-45.

Harvey, Campbell R., and Yan Liu. "Does Scale Impact Skill?." (2016). *Working Paper*

Table 1. Regression of fund performance on lagged fund size using the sample of domestic equity funds from 1981 to 2016

	OLS				S-Estimator			
	Market-Adj	Beta-Adj	3-Factor	4-Factor	Market-Adj	Beta-Adj	3-Factor	4-Factor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LOGTNA	-0.035	-0.034	-0.027	-0.027	-0.016	-0.027	-0.025	-0.018
	(-3.26)	(-3.04)	(-2.43)	(-2.45)	(-1.38)	(-2.27)	(-2.05)	(-1.46)
LOGFAMSIZE	-1.336	-1.472	-1.301	-1.364	-2.100	-3.272	-2.547	-2.012
	(-1.55)	(-1.83)	(-1.74)	(-1.77)	(-1.61)	(-1.86)	(-1.62)	(-1.09)
TURNOVER	0.019	0.024	0.025	0.024	0.061	0.041	0.053	0.050
	(0.99)	(1.13)	(1.14)	(1.13)	(2.38)	(1.48)	(2.10)	(1.86)
AGE	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(2.43)	(2.28)	(2.25)	(2.26)	(1.08)	(1.13)	(1.45)	(0.93)
EXPRATIO	0.026	0.022	0.020	0.020	0.039	0.028	-0.004	0.024
	(0.58)	(0.46)	(0.44)	(0.43)	(0.85)	(0.63)	(-0.07)	(0.50)
TOTLOAD	-0.007	-0.008	-0.008	-0.008	-0.004	-0.006	-0.006	-0.006
	(-1.67)	(-1.99)	(-1.99)	(-1.98)	(-1.09)	(-1.53)	(-1.42)	(-1.46)
FLOW	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.001
	(2.53)	(2.62)	(2.64)	(2.62)	(1.72)	(2.52)	(2.53)	(2.03)
LAGFUNDRET	0.013	0.016	0.016	0.016	0.015	0.019	0.018	0.019
	(2.96)	(3.39)	(3.38)	(3.38)	(3.13)	(3.50)	(3.46)	(3.50)
No. of Months	431	431	431	431	431	431	431	431

Notes: This table presents the Fama-MacBeth (1973) estimates of monthly gross returns (in percentage) on fund characteristics lagged 1 month using OLS. The sample includes domestic equity funds that fall within fund size quantiles two to five. Columns (1)–(4) provide OLS results for regressions as in CHHK, while Columns (5)–(8) report the result of outlier-robust regressions, i.e., S-estimation. Returns are adjusted using the market model (Market-Adj), CAPM (Beta-Adj), and Fama-French three and four factor models. LOGTNA and LOGFAMSIZE are the natural logarithms of fund TNA and one plus the family TNA (less the fund), respectively. TURNOVER is the fund turnover, AGE is the number of years since inception, EXPRATIO is the annual fees and expenses scaled by fund TNA, and TOTLOAD is the sum of front-end, rear-end, and deferred sales charges also scaled by fund TNA. FLOW is the percentage of new fund flow in the fund in the past year. LAGFUNDRET is the buy and hold return of the fund over the prior 12 months. The sample is from 1981 to 2016. The *t*-statistics in parentheses are adjusted for serial correlation using Newey-West (1987) lags of order three.