Distribution Agreement

Stefan Zotovic April 9, 2019

Returns to Scale in the Hedge Fund Industry

By

Stefan Zotovic

John Kim, Ph.D.
Advisor

Economics

John Kim, Ph.D.

Advisor

Samiran Banerjee, Ph.D.

Committee Member

Robert Roth, Ph.D.

Committee Member

2019

Returns to Scale in the Hedge Fund Industry

By

Stefan Zotovic

John Kim, Ph.D.
Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Economics

2019

Abstract

Returns to Scale in the Hedge Fund Industry
By Stefan Zotovic

Financial economists and investors alike have long tried to understand the driving forces behind hedge fund performance. Progress has been made in, for example, distilling manager skill from market and other risk exposures. However, one key question has been left unanswered, or rather, poorly answered: Is the size of a fund informative in predicting its future returns? In the literature, there have been numerous and contradictory attempts to pin down this scale effect. However, econometric bias and uncertainty in the methodologies have interfered with the efforts. Armed with an enhanced recursive demeaning methodology, we find significant evidence of decreasing returns to scale at the fund level which set in after an optimal size. Importantly, we show that it is necessary to account for nonlinearity in the fund size effect. Next, we add to the literature by showing that there are no significant decreasing returns to scale at the industry level. Moreover, the fund-level returns to scale are robust to inclusions of industry size. We reformulate this result as an investment strategy by developing a method to sort funds into those experiencing increasing returns to scale (IRS) and those experiencing decreasing returns to scale (DRS). Then we build a portfolio that is long IRS funds and short DRS funds and show that significant positive returns can be generated by capitalizing on the size-performance relationship.

Returns to Scale in the Hedge Fund Industry

By

Stefan Zotovic

John Kim, Ph.D.
Advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Economics

2019

Contents

1 Introduction

Financial economists and investors alike have long tried to understand the driving forces behind returns generated by actively managed funds. Progress has been made in, for example, distilling manager skill from market and other risk exposures. However, one key question has been left unanswered, or rather, poorly answered: Is the size of a fund informative in predicting its returns?

This question is more relevant now than ever before in the hedge fund industry, which has experienced significant growth since the Great Financial Crisis of 2008. Artificially low interest rates have led yield-seeking investors to search for various ways to deploy their capital. In the third quarter of 2018, the industry reached an all-time high in assets under management (AUM) with just over $3 trillion, excluding funds of funds.[1] Moreover, the number of active funds has grown to over 15,000, accompanying the growth in aggregate industry AUM. With the industry growing by all measures, size emerges as a crucial factor to consider when evaluating hedge fund performance.

Though the returns-to-scale question is at its core a practical one, it has major implications for investment theory as well. Berk and Green (2004) propose a framework for active portfolio management that has decreasing returns to scale as a cornerstone. There are three key elements of the model. First, investors allocate capital to funds competitively. Second, manager skill (i.e. the ability to generate alpha) is heterogeneous; however, there are decreasing returns to scale in deploying manager alpha. Finally, investors learn about manager skill from past performance, and rationally allocate capital to those managers that have shown higher alpha. In model equilibrium, there is no risk-adjusted return. This argument is essentially a no-arbitrage argument, made clear by a consideration of two funds: an active fund and a passive fund. If the active funds generate alpha, investors will rationally allocate capital to the fund until alpha is eroded fully and the value of investing in either fund is the same. This no-arbitrage argument, of course, assumes decreasing returns to scale. If

[1] BarclayHedge

they were not present in deploying the active fund manager's alpha, the fund would grow arbitrarily large with investor flow increasing ad infinitum to the active fund. Therefore, there exists the theoretical prior that there should be decreasing returns to scale at the fund level.

Active money management can be decomposed into two primary industries: the mutual fund industry and the hedge fund industry. While the two types of funds are similar in that they are both actively managed portfolios of investments, they diverge greatly in almost every other aspect. Hedge funds often have much fewer investors and take on higher risks and leverage in investing. They typically require minimum investments, so the target investor is a high net-worth individual or another investment fund. In addition, it is common for them to impose a certain lockup period in which funds cannot be called by the investors. This allows the funds to engage in more illiquid investments. Mutual funds, on the other hand, are forced to stick to more liquid investments, as an investor run on funds would be very problematic otherwise. While both mutual funds and hedge funds charge a fixed management fee, hedge funds have additional performance incentives for the manager. These incentives are subject to high water marks and hurdle rates, to provide some assurance of return to investors. Finally, the mutual fund industry is much more regulated than the hedge fund industry.

The central support for fund-level decreasing returns to scale in the hedge fund industry is the idea that larger funds will eventually run into liquidity issues, and they will have a tougher time beating the market. On the other hand, however, fixed costs of funds act as an opposing force to decreasing returns to scale. As fund size increases, the fixed fees and payments of the fund decrease relatively, so that there are increasing returns to scale. Certainly, there are other attributes of a fund which might lend themselves to increasing, decreasing, or constant returns to scale. The tug-of-war between these opposing effects of scale ultimately determine the scale effect of the fund.

Decreasing returns to scale have been detected empirically in the mutual fund industry at the fund level (Zhu (2018)). Pástor, Stambaugh, and Taylor (2015) failed to detect decreasing returns in the mutual fund industry at the fund level, but Zhu (2018) revisited this result using an enhanced recursive demeaning methodology. In the hedge fund literature, Agarwal, Daniel, and Naik (2004), Fung, Hsieh, Naik, and Ramadorai (2008), Teo (2009), and Joenväärä, Kosowski, and Tolonen (2013) find evidence of decreasing returns to scale at the fund level. However, Cao and Velthuis (2017) find no evidence of decreasing returns to scale at the fund level, using the recursive demeaning methodology of Pástor, Stambaugh, and Taylor (2015). While the attempts prior to Cao and Velthuis (2017) were plagued with bias, their attempt was undermined by uncertainty and weakness in hypothesis testing. Given the mixed findings for returns to scale and the recent advances in financial econometrics, it is important to revisit the question of returns to scale at the fund level in the hedge fund industry.

Similarly, there are mixed findings for the scale effect of industry size, as well as different ways of defining "industry", itself. In our analysis, we use only funds that follow equity-oriented strategies and call the sum of all equity-oriented funds the "industry", like Cao and Velthuis (2017). Specifically, this definition includes funds that are long/short equity, equity market neutral, and event-driven. Getmansky (2012) finds that if investors chase a category of hedge funds with good past performance, then the probability of funds liquidating in that category increases. This suggests that as more funds flock to a category, alpha is eroded and some are forced to liquidate (i.e. decreasing returns to scale at the defined category level). Naik, Ramadorai, and Stromqvist (2007) regress Fung-Hsieh seven-factor model alpha on lagged aggregate sizes of various strategies and find significant negative estimates of the size coefficients for four out of the eight strategies. They define the strategies according to the distinct style factors in hedge funds returns, found by Fung and Hsieh (1997) and Brown and Goetzmann (2001). Two of these strategies are comprised predominately by equity funds, which we use in our analysis. For the strategy named 'Multi-Process' which contains event-driven funds, the estimate for the coefficient is negative and not significant. Moreover, for the strategy named 'Security Selection' (comprised of long/short equity hedge, equity hedge, and equity market neutral funds), the estimate for the coefficient is positive and not significant. This strategy category corresponds closely with our industry definition.

Cao and Velthuis (2017), on the other hand, find that there are decreasing returns to scale at the industry level. They then further break down into strategy levels, corresponding to the three categories that comprise the industry, and find significant decreasing strategy-level returns to scale only for the Long/Short Equity Hedge strategy, but not for Equity Market Neutral or Event-Driven. Note the key differences in how Naik, Ramadorai, and Stromqvist (2007) and Cao and Velthuis (2017) define strategy level. We do not do a strategy-level analysis in addition to the industry-level, as we have defined it, and so we sidestep the naming confusion. It makes the most sense to simply group the equity funds and see if there is an aggregate size effect, as the constraints are that of the entire equity market.

As mentioned, hedge funds are much less regulated than mutual funds. Importantly, hedge funds are not required to report returns periodically. Thus, any data on hedge funds is voluntary. Ironically, though there is a relative lack of regulation on hedge funds, one key regulation motivates funds to report their data. Rule 502(c) of the 1933 Securities Act prohibits hedge funds from publicly advertising themselves; instead, they list their funds in various databases to appeal to investors (Jorion and Schwarz (2012)). The listing process is subject to several biases in the data, including survivorship bias, self-selection bias, and backfill bias. Survivorship bias is associated with the phenomena that only the funds that are able to generate returns are able to last. Self-selection bias is associated with certain funds choosing to report their returns, which may not be representative of all hedge funds. Backfill bias is the tendency for funds to misreport returns for previous months when beginning to report to a database. We mitigate these effects by taking certain measures in our data cleaning. In addition to specific biases in the data, there is the concern with hedge funds that they outright misreport returns. Suggestive of manipulation, Bollen and Pool (2009) find that the distribution of monthly hedge fund returns has a low number of small losses and a large number of small gains, creating a kink point around zero. However, this can be attributed to the incentive fee mechanism that somewhat distorts net returns with respect to gross returns, in addition to the underlying illiquid assets and fixed-income securities hedge funds tend to invest in which have yields minimally bounded at zero (Jorion and Schwarz (2014)). Therefore, we are more at peace with the data, given no direct proof of manipulation in reporting of returns.

The asymmetry in net returns created by the incentive mechanism in hedge funds motivates us to use the gross return measure. In analyzing whether fund size constrains performance, it is more useful to look at what the fund manager literally made, rather than the net return, which takes out the fees. Most funds do not report gross returns to the database, but this measure can be backed out from the given data (Jorion and Schwarz (2014)). Gross returns are driven not just by manager skill, but also by exposures to factor variations. Therefore, we use two risk-adjusted alpha measures, Fung-Hsieh eight-factor model Alpha and Capital Asset Pricing Model Alpha, giving us three performance measures to analyze (Fung and Hsieh (2004)). For fund size, the data provides a measure of assets under management. Our performance-size relationship is compared using these three performance measures and the fund size.

A simple OLS regression of performance on size suffers from an omitted variable bias, because unobserved manager skill is contained within the error term. If skill and fund size are positively correlated, then there is an upward bias in the estimator. In other words, the regression might fail to show decreasing returns to scale even in their presence. In order to control for the unobserved heterogeneity in skill across funds and eliminate the omitted variable bias, a fixed effect model is used. The fixed effect model is generally used to control for unobserved heterogeneity, under the assumption that the heterogeneity is time-invariant. In our case, the unobserved heterogeneity is manager skill, so the assumption is reasonable and the model helps to eliminate the omitted variable bias. Solving requires a demeaning procedure to see if variations in fund size around the mean have an effect on variations in performance around a mean. However, this method suffers from finite sample bias (Hjalmarsson (2010)). Pástor, Stambaugh, and Taylor (2015) build on work by Moon and Phillips (2000), Sul, Phillips, and Choi (2005), and Hjalmarsson (2010) to develop a procedure in which all time-series demeaned variables are substituted with their forward-demeaned counterparts to remove dependencies between the demeaned fund size and error term. In this procedure, used by Pástor, Stambaugh, and Taylor (2015) for mutual funds and by Cao and Velthuis (2017) for hedge funds, a regression is run of forward-demeaned returns on the forward-demeaned fund size, while instrumenting the forward-demeaned fund size by backward demeaned fund size. Both fail to find decreasing returns to scale in their respective industries at the fund level. Zhu (2018) finds that this recursive demeaning procedure suffers from a misspecification resulting from a restriction in not allowing a constant in the first stage of the two-stage least squares regression, which is flawed when dealing with fund sizes. This misspecification increases estimation uncertainty and thereby reduces the power in hypothesis testing. Using an enhanced estimator, Zhu (2018) finds fund-level decreasing returns to scale in the mutual fund industry. We use this enhanced estimator for the hedge fund industry and find that significant decreasing returns to scale eventually set in at the fund level. By including a squared term, we allow for non-linearity in the relationship between size and returns. We find that the estimate for the coefficient on the size term is positive and significant and the estimate for the coefficient on the squared size term is negative and significant. This implies that funds experience increasing returns to scale up to an optimal point after which they experience decreasing returns to scale. We find that it is important to account for non-linearity in the fund size effect, which is uncommon in the literature. Moreover, there are no significant returns to scale at the industry level. Thus, the size of the hedge fund industry does not have a significant impact on the returns of a particular fund. This could be due to the relatively small fraction of the total stock market capitalization that hedge funds make up (about 0.56% monthly average). These results are robust to several different performance measures and industry size specifications.

We analyze the practical importance of identifying funds that are experiencing increasing returns to scale (IRS) and funds that are experiencing decreasing returns to scale (DRS). Intuitively, IRS funds are underfunded and should have a greater ability to generate alpha than DRS funds. We design a strategy that takes a long position in IRS funds and a short position in DRS funds and find that it consistently yields positive returns for all performance measures and with various identification methods for determining IRS/DRS. Once we carefully account for non-linearity in sorting funds experiencing IRS versus funds experiencing DRS, the results are strongest. The portfolio analysis shows that returns to scale have important implications for picking out which funds are underfunded. This is important for investors who want to choose to allocate capital to funds that can generate higher alpha.

The remainder of the paper will proceed as follows: Section 2 will go into more detail about the data used, Section 3 will motivate and outline the enhanced recursive demeaning methodology, Section 4 will discuss results, and Section 5 will conclude. Tables are found in Appendix A and figures are found in Appendix B.

2 Data

The data comes from the Trader Advisor Selection System (TASS), which includes data on over 7,500 active hedge funds and funds of funds and over 11,000 graveyard funds. It includes variables on return, assets under management, performance fees, fund types, date that the fund starting reporting to TASS, etc.

2.1 Data Cleaning

Per Chung and Kang (2016), we filter out funds that do not report monthly returns, funds that report returns not denominated in the US dollar, funds with unknown styles, and funds that report before-fee returns, not after-fee. Following Aggarwal and Jorion (2010), we filter out duplicates in the data. Returns prior to 1994 are not considered because TASS does not adequately cover graveyard funds before this date (Agarwal, V., Daniel, N., and Naik, N. (2004)). To avoid "stale returns", funds that report the same returns for at least three consecutive months are excluded from the data (Joenväärä, Kosowski, and Tolonen (2013)). Funds that do not follow equity-oriented strategies are filtered out, per Cao and Velthuis (2017). This measure is taken to allow for proper industry level measurements, as the value of all securities these funds can invest in is easily trackable (i.e. the stock market). This corresponds to funds with category designations of 'Long/Short Equity Hedge', 'Equity Market Neutral', and 'Event-Driven' in the TASS database.

Hedge fund data is, by nature, prone to certain biases, including survivorship bias, self-selection bias, and backfill bias. Survivorship bias is the bias rooted in focusing on funds that have been able to stay afloat and generate returns, the survivors. The survivorship bias is addressed by including active funds as well as graveyard funds; in other words, both the survivors and non-survivors are included. Self-selection bias is associated with funds choosing to report their returns. The presence of this bias would indicate that the sample is not representative of the world of hedge funds. Fung and Hsieh (1997) offer anecdotal evidence that the magnitude of self-selection bias is insignificant due to offsetting factors at play. On one hand, highly successful funds, such as the George Soros Quantum Fund, have no need to advertise good returns for a chance at more investor capital. On the other hand, managers eager to grow AUM may relish the opportunity to report their returns to databases. However great or small the magnitude of the bias, there is no real way to avoid it as the entire TASS database is derived from hedge funds self-reporting. As a precautionary step, funds that do not report returns for at least 24 periods are filtered out (Chung and Kang (2016)), so that funds which encounter bad returns and stop reporting are not taken into consideration.

Backfill bias is the bias associated with posting earlier returns when a fund begins reporting to a database. The "incubation period" is the period of time in between fund inception and the date at which the fund begins reporting returns. During the incubation period, fund managers use capital of friends and family to attempt to generate good returns. Then, they market their record to database vendors. When the vendors input the data, they need to backfill historical returns (Fung and Hsieh (2000)). TASS reports the dates that define the incubation period, inception date and reporting date. Fung and Hsieh (2000) report that the median incubation period is 343 days, or about 12 months using the TASS database from 1994 to 1998. Jorion and Schwarz (2012) find an average incubation period of 27 months, using the TASS database as well as the HFR database. Given the estimates for the length of the incubation period, it is important to gauge how they impact the returns data. Agarwal and Jorion (2010) define a "non-backfilled" fund as one for which the the incubation period is less than 180 days. The lag of 180 days is chosen as a buffer for funds that may be focused on gathering assets rather than investing initially. They find that the backfill bias is quite substantial; backfilled funds outperform the non-backfilled funds for the first four years. Due to these findings, it is common in the literature to filter out the first portion of a fund's returns.

Chung and Kang (2016) choose to filter out the first 18 months of returns for each fund. Joenväärä, Kosowski, and Tolonen (2013), Teo (2009), and Zhong (2008) exclude the first 12 months of returns. Fung, Hsieh, Naik, and Teo (2016) exclude the first 24 months of returns. However, Aggarwal and Jorion (2010) emphasize that this filter is rather arbitrary. The literature broadly chooses to eliminate at least 12 months based on the median backfill period of 343 days found by Fung and Hsieh (2000). Even with this estimate, however, about half of the funds that backfill will still have included some returns from the incubation period. On the upper end of the spectrum, choosing to eliminate 24 months of returns for a fund is excessive when considering funds with no backfill.

This bias is especially problematic for the question of returns to scale. Funds tend to be smaller at inception, before they have had time to generate returns and grow the assets under management. Backfilling higher returns early on, and then reporting actual returns later on could show decreasing returns to scale when there are not any present. Therefore, it is important to address this problem in a conservative and comprehensive way. We decide to filter out backfilled returns by excluding all returns for a fund before the date it was added to the database, per Aggarwal and Jorion (2010).

2.2 Size Proxies

Fund size is adjusted for inflation using the Consumer Price Index (CPI) to January 2000 dollars. Though Cao and Velthuis (2017) only use fund size, we use logarithm of fund size as a second size proxy for comparison, which is used widely in both mutual fund and hedge fund data (see, e.g., Pástor, Stambaugh, and Taylor (2015), Zhu (2018), Chen, Hong, Huang, and Kubik (2004), Ferreira, Keswani, Miguel, and Ramos (2013), Yan (2008), Ammann and Moerth (2005), Naik, Ramadorai, and Stromqvist (2007), Teo (2009)).

We define the industry level in two ways for robustness. First, we use the aggregate of the CPI-adjusted fund AUM in dollars. Next, we use the previous measure divided by the total U.S. stock market capitalization (Cao and Velthuis (2017)). Therefore, the latter measure looks at the hedge fund industry as a function of the total stock market, which might be a better measure for capturing capacity constraints and crowding.

2.3 Deriving Performance Measures: Gross Return

There are several measures of performance used in the literature for hedge funds. One is the net-of-fee return, which is most common in the data (as mentioned in Section 2.1, we filter for this measure in the data). As the name suggests, this return measure is taken after management and incentive fees are subtracted. Typical performance fees are around 20% of profits, and typical management fees are around 2% of assets under management (both on a per annum basis). Hedge funds also generally have high water marks and hurdle rates. A high water mark is a previous high value achieved by the fund, with performance fees kicking in after the high water mark is surpassed. For example, if a hedge fund with $100 generates

20% returns and reaches $120, but then loses $40, incentive fees will only kick in after the managers make back the lost money, i.e. above $120. The hurdle rate is the minimum rate of return expected, above which managers are given incentive fees.

Net-of-fee returns are asymmetrical by nature, due to the performance incentives. Take for example a fund that does not charge any management fee but does charge a 2% performance fee. Ignoring high water marks and hurdle rates, if it has a return of 10%, then the reported net-of-fee return is 8%. However, if it has a return of -10%, then that is the net-of-fee return that it reports, as the performance incentive is not met. Thus, the positive returns are compressed towards zero (Jorion and Schwarz (2014)). Though this example is extreme, it is illustrative of the asymmetry.

In order to avoid the distortions and asymmetries of the net-of-fee return measure, it is important to use the gross return measure. Unfortunately, this measure is generally not included in the data. However, it can be backed out from the given data, using net-of-fee return, net asset value (NAV), management fee, performance fee, and high water mark. Although the TASS database does not report incentive fee accrual period, Jorion and Schwarz (2014) find that the distribution of gross returns for assuming quarterly accrual period and assuming annual accrual period is nearly identical. Like them, we report results assuming an annual accrual period.

A nuance to consider is that the net-of-fee returns can be different for all investors (who do not invest capital in the same starting time). This is due to the incentive mechanisms, which are particular to the specific investor as a result of the high water mark. Following Jorion and Schwarz (2014), we assume that all funds have a high water mark and that all investors are invested at the previous high water mark. This is like assuming that all investors enter a fund at inception and no new investors allocate capital afterwards. At first, this seems like an unreasonable assumption. For this reason, Cao and Velthuis (2017) outline a very complex procedure for backing out gross returns that, among other features, has a first-in-first-out (FIFO) rule to determine which investor's capital is leaving or entering the fund. However, according to Jorion and Schwarz (2014), TASS reports net-of-fee returns for a representative share class of investors who all have the same high water mark, which is usually the oldest. Therefore, our assumption is reasonable and fits the data reporting more accurately. Similarly, Teo (2009) assume that returns accrue to a first-year investor in calculating gross returns.

Per Jorion and Schwarz (2014), the gross return is backed out as follows. First, the high water mark (HWM) at period t is tracked as the maximum value for net-asset-value (NAV) up to period $t-1$. Then the accrued incentive fee, IF_t is calculated as

$$IF_t = \frac{f}{1-f} MAX(NAV_t - HWM, 0) \tag{1}$$

where f is the incentive fee percentage. Then the management fee, MF_t is calcuated as

$$MF_t = \frac{m}{1-m}(NAV_t + IF_t) \tag{2}$$

where m is the management fee percentage. Finally, the gross return, R_t is computed as

$$R_t = \frac{NAV_t - NAV_{t-1} + MF_t + IF_t - IF_{t-1}}{NAV_{t-1} + IF_{t-1}} \tag{3}$$

2.4 Deriving Performance Measures: Risk-Adjusted Alphas

In evaluating the return of a fund, it is crucial to distinguish between the portion of return that comes from general risk exposures of the fund and the portion that comes from manager skill, or alpha. The traditional view for investment vehicles, as expressed by the Capital Asset Pricing Model (CAPM), is that all returns not related to the market are attributable to alpha. However, hedge funds are very unique in the investment industry with regards to the loose restrictions they face in allocating capital, often engaging in short-selling, leveraging, and derivatives. Much research focused on hedge-fund-specific risk exposures has identified key factors that can be used to generate a more comprehensive and appropriate alpha measure. Therefore, in addition to CAPM alpha, we use Fung-Hsieh alpha (FH alpha) based on the Fung and Hsieh (2004) eight-factor model. The eight factors used to calculate FH alpha are S&P 500 return minus risk free rate (MKT), Wilshire small-cap minus large-cap return (SZSPRD), yield spread of the U.S. 10-year Treasury bond over the three-month Treasury bill, adjusted for the duration of the U.S. 10-year Treasury bond (BNDMKT), change in the credit spread of Moody's BAA bond over 10-year Treasury bond adjusted for duration (CRD-SPRD), excess returns on portfolios of lookback straddle options on currencies (PTFSFX), bonds (PTFSBD), and commodities (PTFSCOM), and Emerging Market Index total return (MSCIEM) (Fung and Hsieh (2001)).[2] These eight factors have great explanatory power for hedge fund returns (Teo (2009)). For CAPM alpha, the only factor is excess market return (MKTRF) (Fama and French (1993)). Excess market return is calculated as the difference between the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) and the one-month Treasury bill rate (from Ibbotson Associates).[3]

In order to derive the alpha measures, excess gross return (gross return minus risk-free rate) is regressed on the factors (or single factor in the case of CAPM alpha) using rolling regressions, per Agarwal, Green, and Ren (2017). Using rolling regressions allows us to have a unique alpha for a fund at time t, not simply a time-invariant fund alpha, as is used by Cao and Velthuis (2017). The former uses one regression per fund per month, whereas the latter uses one regression per fund. The use of rolling regressions better captures the risk-adjusted performance for a fund at a specific time. The months $t-23$ to t are used in the following regression to determine FH alpha factor loadings, where the sum of all factors at time t multiplied by their coefficients is represented by the sum notation:

$$R_{i,t} - R_{rf,t} = \alpha_i + \sum_{factors} \beta_{i,factor} FACTOR_t + u_{i,t} \tag{4}$$

[2]David A. Hsieh's Data Library, http://faculty.fuqua.duke.edu/ dah7/DataLibrary/TF-FAC.xls
[3]Fama and French Data Library, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data$_library.html$

Similarly, the months $t - 23$ to t are used to determine the CAPM alpha factor loading:

$$R_{i,t} - R_{rf,t} = \alpha + \beta_{i,mktrf} MKTRF_t + u_{it} \tag{5}$$

Next, the alpha measures are calculated by subtracting excess actual gross return from the model-fitted return in each case at time t. As a result of the window size for the regressions, alpha measures for each fund exist starting for each month after the twenty-third in the sample.

3 Methodology

Let $R_{i,t}$ be the return for fund i at time t. Let $x_{i,t}$ be the fund size for fund i at time t . Note that lagged fund size is of interest for seeing the effect of fund size on fund performance; in other words, for fund size, the period $t - 1$ is used instead of t. This is because the fund size at the beginning of an investment period is relevant for performance of a fund at the end of the investment period.

3.1 OLS Regression

In order to evaluate the scale effect, much of the literature uses a pooled OLS regression, which is simply an OLS regression run on panel data (see, e.g., Chen, Hong, Huang, and Kubik (2004), Ferreira, Keswani, Miguel, and Ramos (2013), Goetzmann, Ingersoll, and Ross (2003), Yan (2008)). Such an OLS regression of returns on lagged fund size would take the form:

$$R_{i,t} = \alpha + \beta x_{i,t-1} + u_{i,t} \tag{6}$$

Note that skill is an unobservable variable that is consequently found in the error term $u_{i,t}$. If the fund size were random and independent from manager skill, then the OLS estimate for β , $\widehat{\beta}$, in equation (6) would reveal the returns to scale. If $\widehat{\beta} < 0$, there are decreasing returns to scale. Similarly, $\widehat{\beta} = 0$ indicates constant returns to scale and $\widehat{\beta} > 0$ indicates increasing returns to scale. The latter two cases imply that the investment strategies are infinitely scalable, as incoming funds have no diminishing effect on the returns. In such a scenario, there would always be a net present value for an investor.

Unfortunately, skill and size are not independent, which leads to a bias in $\widehat{\beta}$ in the simple OLS regression. Larger funds might have the capability to hire more competent managers. Similarly, more skilled managers may be able to generate higher return and attract investor capital, thereby increasing fund size. This results in a non-zero expected value of the error term conditional on lagged fund size, as formulated in equation (7), which implies an endogeneity problem.

$$E(u_{i,t}|x_{i,t-1}) \neq 0 \tag{7}$$

Moreover, there is an upward bias in $\widehat{\beta}$ because skill has a positive effect on performance and size and skill are likely positively correlated (Angrist and Pischke (2009)). Therefore, the estimator may show constant or increasing returns to scale when, in fact, there are decreasing returns to scale.

3.2 Fund Fixed Effects

When regressing returns on size, there are two potential types of variation that can be taken into consideration: inter-fund variation and intra-fund variation. Inter-fund variation is the the variation in average returns from one fund to another. Intra-fund variation is the variation in returns over time within a fund. The OLS regression in model (6) focuses on the inter-fund variation, because it assumes that there is no cross-sectional variation in fund skill. This implies that all funds have the same α. Given that there is certain heterogeneity in manager skill, this is an unreasonable assumption. In order to control for the unobserved heterogeneity in skill across funds, and, therefore, eliminate the endogeneity problem, a fixed effect model is used. The fixed effect model holds fixed the average effects of each fund, as shown below.

$$R_{i,t} = \alpha_i + \beta x_{i,t-1} + u_{i,t} \tag{8}$$

In the OLS regression model, the intercept is the same for all funds, simply α. However, in the fixed effect model there is a fund-specific intercept, α_i. Contrary to the inadvertent focus on inter-fund variation of the OLS panel regression, this model isolates the intra-fund variation. Specifically, it tells us how the variations in fund sizes around their means relates to variation in returns around their means. Fixed effect models are relevant under the assumption that the unobservable factors that lead to endogeneity are time-invariant. In our case, it is a relatively innocuous assumption that manager skill is time-invariant. Each fund then has a fixed level of skill, given by α_i, which absorbs any variation in performance due to cross-sectional variation in fund skill.

The model in (8) is consistent with the model of Berk and Green (2004), which is outlined in Section 1. The Berk and Green (2004) model assumes that there is heterogeneity in manager skill among funds, but that manager skill of a particular fund is constant through time. Investors perceive a level of manager skill that may differ from the true skill, and they allocate capital to managers who they perceive have higher alpha. The relationship between perceived skill and true time-invariant skill is summarized in the following equation (Pástor, Stambaugh, and Taylor (2015)):

$$perceived_{i,t} = true_i + noise_{i,t} \tag{9}$$

As the perceived alpha varies, there are variations in the fund size, with investors rationally allocating capital based on which funds they believe will generate higher returns. Because the perceived skill is generally not equal to the true skill, the fund size is generally sub-optimal. In the framework, this causes the true expected return to change based on diseconomies of scale. For example, when a fund's perceived alpha exceeds the true alpha, the fund exceeds optimal size, and the true expected future return is lower given decreasing returns to scale. On the other hand, when a fund's perceived alpha is lower than the true alpha, the fund size is lower than optimal, and the true expected return is higher. Therefore, the fixed effect model in (8) corresponds to the theoretical framework of Berk and Green (2004) with constant fund manager skill, and implies that there are decreasing returns to scale whenever true alpha of a fund decreases with size.

Though the fixed fund effects model corrects for the omitted variable bias, it introduces a new finite-sample bias because of the demeaning method used in estimating fixed effect models. To illustrate this bias, first consider the value for the estimator of β in model (6):

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_{i,t-1} - \overline{x}_i)(R_{i,t} - \overline{R}_i)}{\sum_{i=1}^{n}(x_{i,t-1} - \overline{x}_i)^2} \tag{10}$$

where

$$\overline{x}_i = \left(\sum_{t=1}^{T_i} x_{i,t}\right)/n \tag{11}$$

and

$$\overline{R}_i = \left(\sum_{t=1}^{T_i} R_{i,t}\right)/n \tag{12}$$

are the sample means of the x_i and R_i respectively. The notation implies n funds in the sample and T_i time periods in the sample for fund i. Now consider the demeaned variables:

$$\widetilde{x}_{i,t-1} = x_{i,t-1} - \overline{x}_i \tag{13}$$

$$\widetilde{R}_{i,t} = R_{i,t} - \overline{R}_i \tag{14}$$

$$\widetilde{u}_{i,t} = u_{i,t} - \overline{u}_i \tag{15}$$

where $\overline{x}_i$ and $\overline{R}_i$ are defined as before, and

$$\overline{u}_i = \left(\sum_{t=1}^{T_i} u_{i,t} \right) / n \tag{16}$$

These three variables are demeaned by the full-sample time series means. Then, consider the demeaned model of (8):

$$\widetilde{R}_{i,t} = \beta \widetilde{x}_{i,t-1} + \widetilde{u}_{i,t} \tag{17}$$

Note that there is no intercept in the demeaned model (17), and that the value of the estimator is the same in both the demeaned model (17) and the fixed-effects model (8). Using the demeaned variables, (10) can be rewritten as:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} \widetilde{x}_{i,t-1} \widetilde{R}_{i,t}}{\sum_{i=1}^{n} \widetilde{x}_{i,t-1}^2} \tag{18}$$

In order to investigate bias as the difference between β and $\widehat{\beta}$, (18) can be rewritten as:

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n} \widetilde{x}_{i,t-1} \widetilde{u}_{i,t}}{\sum_{i=1}^{n} \widetilde{x}_{i,t-1}^2} \tag{19}$$

Therefore, the estimator $\widehat{\beta}$ is unbiased if the numerator of the second term in (19) is equal to zero. This holds if $\widetilde{x}_{i,t-1}$ and $\widetilde{u}_{i,t}$ are not correlated. However, from (13), we can see that the demeaned fund size depends on all values of fund size across all t, not just periods before $t-1$. The future values of fund size are certainly affected by the error term in time t. Specifically, $\widetilde{x}_{i,t-1}$ is negatively correlated with the innovation in $x_{i,t}$ which is positively correlated with $u_{i,t}$. This implies a negative correlation between $\widetilde{x}_{it-1}$ and $\widetilde{u}_{it}$.

Intuitively, the demeaning process makes each specific value of fund size at a time t in the sample dependent on values of fund size for all t in the sample. For example, higher fund sizes in the future would increase the mean, and thereby decrease the values of the earlier demeaned fund sizes. This leads to a downward bias in the estimator, which may predict decreasing returns to scale when there are none.

3.3 Recursive Demeaning

Demeaning the variables by the full time-series means leads to a biased estimator because of correlation between demeaned size and demeaned innovation. Hjalmarsson (2010) develops a procedure in which all demeaned variables are replaced with their forward-demeaned counterparts, building on work by Moon and Phillips (2000) and Sul, Phillips, and Choi (2005). This procedure is adapted to the economies of scale problem by Pástor, Stambaugh, and Taylor (2015), and subsequently by Cao and Velthuis (2017), by estimating a two-stage least-squares (2SLS) regression of recursively demeaned variables. To outline the recursive demeaning methodology, first, define the forward demeaned variables:

$$\overline{R}_{i,t} = R_{i,t} - \frac{1}{T_i - t + 1}\sum_{s=t}^{T_i} R_{i,s} \tag{20}$$

$$\overline{x}_{i,t-1} = x_{i,t-1} - \frac{1}{T_i - t + 1}\sum_{s=t}^{T_i} x_{i,s-1} \tag{21}$$

$$\overline{u}_{i,t} = u_{i,t} - \frac{1}{T_i - t = 1}\sum_{s=t}^{T_i} u_{i,s} \tag{22}$$

Pástor, Stambaugh, and Taylor (2015) and Cao and Velthuis (2017) use the backward demeaned variable for fund size as an instrument:

$$\underline{x}_{i,t-1} = x_{i,t-1} - \frac{1}{t-1}\sum_{s=1}^{t-1} x_{i,s-1} \tag{23}$$

They estimate the 2SLS regression by first regressing $\overline{x}_{it-1}$ on $\underline{x}_{it-1}$, and then regressing $\overline{R}_{it}$ on the fitted-values from the first stage. Importantly, a zero intercept is imposed in both stages, as shown below.

$$\textit{First stage:} \qquad \overline{x}_{i,t-1} = \rho \underline{x}_{i,t-1} + v_{i,t-1} \tag{24}$$

$$\textit{Second stage:} \qquad \overline{R}_{i,t} = \beta \widehat{\overline{x}}_{i,t-1} + u_{i,t} \tag{25}$$

We refer to this recursive demeaning procedure as RD1 for reference. Using RD1 to estimate the scale effect in β, both Pástor, Stambaugh, and Taylor (2015) and Cao and Velthuis (2017) find no significant effects at the fund level in the mutual fund industry and hedge fund industry, respectively. However, Zhu (2018) argues that this is because this estimator introduces uncertainty that weakens hypothesis testing.

Zhu (2018) runs simulations and calculates the root mean square errors (RMSE) of the various estimators discussed: the OLS estimator, the fixed effects estimator, and the RD1 estimator. RMSE is calculated as $\sqrt{bias^2 + sd^2}$, and is preferred to bias or standard deviation alone in evaluating an estimator. The RD1 estimator generally has a lower RMSE than the OLS estimator, but it has a larger RMSE than the fixed effect estimator in most simulations. Therefore, it is not clear that the recursive demeaning methodology of Pástor, Stambaugh, and Taylor (2015) and Cao and Velthuis (2017) is worthwhile in correcting the finite-sample bias of the fixed-effect model.

In order to enhance the recursive demeaning estimator, two modifications are proposed. First, an intercept is included in the first stage of the regression. Second, pure fund size, $x_{i,t-1}$, is used as the instrument instead of backward demeaned fund size, $\underline{x}_{i,t-1}$. Imposing a zero intercept in the second stage of the regression is necessary because of the fund fixed effect being removed in the recursively demeaned model. However, imposing a zero intercept in the first stage is a clear model misspecification, as it implies that fund size fluctuates around a constant mean. In turn, it does not allow the regression much flexibility in finding the best fit and causes an increase in uncertainty in the estimator.

The revised instrument, $x_{i,t-1}$, meets the conditions of relevance and exclusion of Roberts and Whited (2013). The instrument satisfies the relevance condition because it is significantly related to the forward demeaned variable, $\overline{x}_{i,t-1}$, in the first stage of the regression, as both are related to fund size. It satisfies the exclusion condition because the forward demeaned values of error are unlikely to be related at all to fund size in the preceding period $t-1$:

$$E(\overline{u}_{i,t}|x_{i,t-1}) = 0 \tag{26}$$

The updated 2SLS regression takes the following form:

$$\textit{First stage:} \qquad \overline{x}_{i,t-1} = \Theta + \rho x_{i,t-1} + v_{i,t-1} \tag{27}$$

$$\textit{Second stage:} \qquad \overline{R}_{i,t} = \beta \widehat{\overline{x}}_{i,t-1} + u_{i,t} \tag{28}$$

We refer to this updated recursive demeaning methodology as RD2. In the first stage, the demeaned variable for lagged fund size is regressed on the instrument of fund size. An intercept, Θ, is included in this stage. In the second stage, fund performance is regressed on the fitted values for forward demeaned fund size from the first stage. An intercept is not included in this stage so as to be consistent with the demeaned fund fixed effects model. The estimator from this second stage gives an unbiased scale effect. Moreover, in the simulations by Zhu (2018), it has lower RMSE than any of the other estimators, including half the RMSE of the fixed effects estimator. Most of this improvement is due to the better model specification of RD2 in allowing for an intercept in the first stage.

3.4 Characterizing Funds as IRS or DRS

In Section 4.3, we motivate a potential investment strategy that monetizes relationship between size and return. Here, we introduce how the data is collected for the strategy development. In order to determine which funds are experiencing increasing returns to scale (IRS funds) and which funds are experiencing decreasing returns to scale (DRS funds), another rolling regression is performed, this time with a fixed starting month per fund, t_{start} and an end date t that increases by one month for each regression. We perform this procedure for the three performance measures. When using the two alpha measures, t_{start} is the twenty-fourth month for each fund because that is the earliest month for which an alpha measure is available, as a result of the initial rolling procedure. The rolling begins with a minimum window of 12 months, and then increases to 13 and and 14 and so on. The months t_{start} to t are used in following regression:

$$R_{i,t} = \alpha + \beta x_{i,t-1} + u_{i,t} \tag{29}$$

where the return on the left-hand side of the equation is one of the three performance measures, and where $x_{i,t-1}$ is the lagged size. Note that the recursive demeaning procedure developed in the previous section is not used, as only one fund is considered per regression. The sign of the estimate for β gives the result: negative implies DRS and positive implies IRS.

4 Empirical Results

This section gives the empirical results investigating whether or not there are decreasing returns to scale at fund level, while controlling for industry level. Then, we derive a potential investment strategy capitalizing on the scale effect of non-optimally funded funds.

4.1 Fund-Level Returns to Scale

As a first step, we look at the size-performance relationship at the fund level using our enhanced recursive demeaning procedure (RD2) and compare it to the results using the recursive demeaning procedure of Pástor, Stambaugh, and Taylor (2015), used by Cao ad

Velthuis (2017) for hedge funds (RD1). We look at the relationship using linear functional form, per Cao and Velthuis (2017), as well as loglinear function form. In order to account for cross-sectional correlations in between funds belonging to the same TASS 'primary category', we cluster standard errors by category x month. We also cluster by fund because recursive demeaning can potentially cause serial correlation within funds. The results are reported in Table 3. Panel A gives the results for models RD1 and RD2 using linear functional form, while Panel B gives the results for the models using loglinear functional form. The RD1 procedure does not give significant decreasing returns to scale for linear function form, which is consistent with the results of Cao and Velthuis (2017). However, the updated methodology does give significant decreasing returns to scale, with an estimate for coefficient on size of -6.53e-09 and corresponding t-statistic of -2.55 and p-value 0.011. In the first stage of the regression, the estimate for the constant is -1.67e07 with a t-statistic of -6.88, compared to the zero intercept in RD1. RD2 leads to a better first-stage fitting, with an R-squared value of 3.48% as opposed to 0.50% of RD1. Moreover, as a test of instrument strength, RD2 gives a first-stage F-statistic of 24.78, which is much higher than the first-stage F-statistic of RD1, 3.47. Overall, this offers overwhelming support for the inclusion of the intercept in the first stage of the model and the revised instrument.

When taking the logarithm of fund size, the results have a rhyme with the linear functional form. Model RD1 gives insignificant increasing returns to scale, with an estimate for the coefficient on logarithm of size of 0.760, with corresponding t-statistic of 0.19 and p-value of 0.848. This result is reversed using RD2 with a negative estimate for the coefficient on logarithm of size, though it is likewise not significant. RD2 has a much better first-stage fit with an R-squared of 7.26% compared to 0.02% of RD1. Moreover, the first-stage F-statistic for RD2 of 90.71 is significantly greater than the value of 0.78 for RD1.

The results of Table 3 motivate two main conclusions. The primary result is that of the better performance of RD2 over RD1, pertaining to the two better model specifications: including a first-stage intercept and using fund size as an instrument instead of backward demeaned fund size. Secondly, there is not consistency in the results for both functional forms using RD2. While there are significant decreasing returns to scale with linear functional form, the significance disappears with loglinear functional form (though the sign remains negative). However, we claim that it is possible to reconcile the disparity by including squared logarithm of fund size, thereby incorporating non-linearity.

In working with hedge fund data, Cao and Velthuis (2017) only report results using the linear functional form. On the other hand, both Pástor, Stambaugh, and Taylor (2015) and Zhu (2018) report results using both linear and loglinear functional forms for the mutual funds. Moreover, the logarithm of size is used commonly in the investment management literature (see, e.g., Chen, Hong, Huang, and Kubik (2004), Ferreira, Keswani, Miguel, and Ramos (2013), Yan (2008)). It is used so broadly in order to account for a skewed distribution in fund size. Under the logarithm transformation, a given percentage change in size has the same effect on performance for both a small and a large fund. Though we recognize the importance of normalizing the fund size distribution, we claim that the logarithm transformation alone may be insufficient for hedge fund data.

Hedge funds have much more heterogeneity in size and skill than mutual funds, and simply taking the logarithm might mute some of the effects. Using fund size (without the logarithm transformation), we have the prior that there are decreasing returns to scale, given in Table 3. It is possible that using the logarithm transformation gives more gravity to the scale effect of small funds, and, thereby, erases the overall effect of decreasing returns to scale. Therefore, we find it important to include a squared term for logarithm of fund size to allow for non-linearity. By including a squared term, we find a balance with applying the logarithm transformation to the skewed size data without muting some important effects. This nonlinear model specification has precedent in the hedge fund literature (see, e.g., Ammann and Moerth (2005), Naik, Ramadorai, and Stromqvist (2007), Teo (2009), Getmansky (2012)).

Table 4 gives the results for the nonlinear model that includes a squared logarithm of size term. When the only regressor is the logarithm of size, the coefficient is negative but insignificant, as discussed above. However, augmenting the model with the squared logarithm of size yields two significant estimates. The estimate for the coefficient on logarithm of size is 3.665, with a t-statistic of 2.24 and p-value of 0.025, and the estimate for the coefficient on squared logarithm of size is -0.116, with a t-statistic of -2.47 and p-value of 0.013. Importantly, the first coefficient is positive and the second coefficient is negative, which gives the relationship a negative parabolic shape. The implication is that funds experience increasing returns to scale for some time until they reach an optimal size, at which point they begin experiencing decreasing returns to scale. This is consistent with the Berk and Green (2004) model which assumes that funds experience decreasing returns to scale in equilibrium. The shape is illustrated in Figure 1 in Appendix B.

The optimal size can be calculated as the maximum of the following function:

$$y = bx + cx^2 \tag{30}$$

where y is return, x is logarithm of fund size, and b and c are the estimates from the regression. Taking the derivative of the function and setting equal to zero gives the optimal fund size:

$$x^* = -\frac{b}{2c} \tag{31}$$

Plugging in the estimates from Table 4, we get $x^* = 15.8279$. Then, of the 147,153 observations used, we find that 24,722 are to the left of x^* and 122,381 are to the right of x^*. Therefore, most of the funds are in equilibrium with decreasing returns to scale.

Gross return is a measure of the return before the managerial and incentive fees kick in. The alpha measures give the portion of this return that is not due to certain risk exposures. In other words, it is a proxy for the portion of gross return this is attributable to the skill of the fund manager. For example, if a fund generates 1% returns in a given month, it is possible that 90% of that move is due to an auspicious upswing in the S&P 500 and only

10% of it is due to some managerial skill. In essence, this 10% is what the alpha measures are designed to extract. Therefore, it is important to take into consideration alpha measures when investigating the scale effect, in addition to gross return.

The results using gross return are supported by the results using the other performance measures, FH alpha and CAPM alpha, given in Tables 5 and 6 respectively. For both alpha measures, there is a significant positive estimate for the coefficient on the logarithm of size when it is the only regressor, indicating increasing returns to scale. However, when allowing for non-linearity with the quadratic regression, there is a positive and highly significant estimate for the coefficient on logarithm of size and a negative and highly significant estimate for the coefficient on squared logarithm of size. Specifically, in the case of FH alpha, the estimate for the coefficient on logarithm of size is 6.161 with t-statistic of 5.10 and the estimate for the coefficient on squared logarithm of size is -0.161 with t-statistic of -4.89. In the case of CAPM alpha, the estimate for the coefficient on logarithm of size is 8.546 with t-statistic of 5.10 and the estimate for the coefficient on squared logarithm of size is -0.229 with t-statistic of -4.89. Thus, there is a negative parabolic shape to the size-performance relationship for the alpha measures as well.

4.2 Industry-Level Returns to Scale

Next, we look at the impact of industry size on fund performance. Tables 4, 5, and 6 give the results with gross return, FH alpha, and CAPM alpha, respectively. Two industry size measures are included for robustness in the last two columns of each table. In the regressions with industry size, the recursive demeaning is, of course, still used. However, only the forward demeaned fund size is instrumented, not the industry size. Two conclusions are clear and common to all performance measures. First, all of the fund-level results discussed previously hold when controlling for industry size. Second, there are no significant results for industry-level returns to scale. Moreover, there is no consensus in the performance measures for a negative coefficient on industry size; while there are negative estimates when using gross return and FH alpha, there is a positive estimate when using CAPM alpha.

4.3 Scale-Based Investment Strategy

At the fund-level, the evidence shows that there is a negative parabolic size-performance relationship. That is, funds experience increasing returns to scale until equilibrium, at which point they begin to experience decreasing returns to scale. The natural implication is that funds experiencing increasing returns to scale (IRS) are underfunded. From the standpoint of an investor, one of the factors to take into consideration when investing funds is the amount of erosion on return that the capital allocation will cause. All else being equal, it is rational to invest in a fund where there is momentum in the returns with every dollar allocated, i.e. funds which are experiencing increasing returns to scale. Thus, investors could derive profit by identifying which funds are IRS and which funds are DRS at a point in time, and investing capital in those funds that are IRS. In this section, we test this prediction.

We seek to establish a method for characterizing funds as either IRS or DRS, and then back-testing the investment strategy which goes long on IRS funds and short on DRS funds. We test two different methods of identification. The first one (Method 1) is described in Section 3.4, with the sole regressor being lagged logarithm of fund size. We perform rolling regressions with a fixed starting month per fund, t_{start} and an end date t that increases by one month for each regression. Thus, at each regression end date t, we estimate a coefficient for the returns to scale at that time using the sample that starts at the beginning date. Note that it is a simple OLS regression and that the methodologies discussed in Section 3 that account for biases and uncertainties are not required as the regression only includes observations from a single fund. A negative estimate for β implies a DRS fund and a positive one implies an IRS fund. A dummy variable is used to keep track of the identifications.

After characterizing funds as IRS or DRS, we are ready to build a portfolio that capitalizes on the implications. If a fund was experiencing IRS in period $t-1$, we want to have a long position on it in period t. Similarly, if a fund was experiencing DRS in period $t-1$, we want to have a short position on it in period t. Thus, the dummy variable for IRS/DRS is lagged so that for each period t, we have an indicator of whether the fund was experiencing IRS or DRS in period $t-1$. Then, for each month t we build a value-weighted portfolio of IRS funds and a value-weighted portfolio of DRS funds, so that the returns are weighted by the relative size of the funds. To achieve the desired long/short portfolio, we subtract the returns of the DRS portfolio from the returns of the IRS portfolio. This measure is then regressed on a constant to measure average return and significance.

The results for the long/short strategy using Method 1 are given in Panel A of Table 7. The number of observations is given, which is the number of fund/months. For FH alpha and CAPM alpha, there are a total of 97,507 data points, which is slightly less than the number of data points for the previous regressions of 112,088 found in Tables 5 and 6. This is due to the minimum window size of 12 in the rolling regressions. Because of this window, the first 11 months of each fund do not have a corresponding characterization as IRS or DRS. These 97,507 observations correspond to 244 months, for which the value-weighted returns are calculated for IRS and DRS funds. After subtracting the returns in order to simulate the long/short strategy, the result is regressed on a constant to test for significance. Using all three performance measures, there are positive average returns. However, these returns are only significant using CAPM alpha.

Method 2 corresponds to our revised nonlinear functional form established in Section 4.1. We augment the rolling regression used in Method 1 from equation (29) with squared logarithm of fund size, in order to account for the better model specification. In this case, the scale effect is not given simply by the coefficient on logarithm of fund size, like in the Method 1. Rather, it is a function of size. We solve for the scale effect at time t by multiplying the estimate of the coefficient for squared logarithm of fund size by the actual value for logarithm of fund size from the data, and then adding that to the estimate of the coefficient for logarithm of fund size. This scale effect calculation is obtained by taking the derivative of the quadratic model. Like before, we set a dummy variable to hold the information on whether the resulting effect is negative or positive, implying DRS and IRS, respectively. After lagging the variable, we set up the value-weighted long/short portfolio.

The results for Method 2 are given in Panel B of Table 7. All of the returns are positive and higher than the corresponding returns of Method 1. Moreover, they are significant; though the average returns increase, their standard deviations decrease slightly. The strategy yields average monthly returns of 0.145% in excess of the market return for CAPM alpha. Though the results are significant using both FH alpha and CAPM alpha, they are not significant with gross return. Gross returns are driven not just by skills, but also by exposures to factor variations. This reinforces the importance of focusing on proper measures of skill.

We have two central conclusions from the portfolio analysis. First, and more importantly, we show that there is an arbitrage opportunity available by identifying which funds are experiencing increasing returns to scale and which funds are experiencing decreasing returns to scale. Positive returns are generated by taking a long position in funds identified as IRS and taking a short position in funds identified as DRS. Moreover, these positive average returns are robust to different performance measures and different identification methods. Second, we see that Method 2, which incorporates nonlinear functional form, yields consistently better returns than its linear counterpart. That is, it seems to better capture the scale effect. This offers support to our argument in Section 4.1 that the nonlinear functional form is more appropriate for the study of economies of scale in hedge funds.

5 Conclusion

This paper empirically addresses one of the cornerstones of the theoretical framework for active management: decreasing returns to scale. In the literature, there have been numerous and contradictory attempts to pin down the scale effect. However, econometric bias and uncertainty in the methodologies have interfered with the efforts. We revisit the result of Cao and Velthuis (2017) that there are no fund-level decreasing returns to scale, but that decreasing returns to scale exist at the industry level. Armed with an enhanced recursive demeaning methodology, we show that there are, in fact, significant decreasing returns to scale at the fund level which set in after an optimal size. Using a quadratic model, we find that there is a significant positive estimate for the coefficient on logarithm of size and a significant negative estimate for the coefficient on squared logarithm of size, yielding a negative parabolic shape to the size-performance relationship. This result holds when using gross return, as well as risk-adjusted performance measures FH alpha and CAPM alpha. Intuitively, a small fund may experience increasing returns to scale due to a relative decrease in fixed fees necessary to manage and run the fund. Most hedge funds nowadays rely on an organization of human and technological capital to be competitive, so initially, a fund may greatly benefit from increasing in size and spreading this fixed cost. However, eventually, the fund reaches a point where it cannot deploy investment strategies without some erosion in return due to liquidity constraints. At this point, it is in equilibrium, experiencing decreasing returns to scale.

Next, we add to the literature by showing that there are no significant decreasing returns to scale at the industry level. Moreover, the fund-level returns to scale are robust to inclusions of industry size in the multivariate regression. For equity funds, this could be due to the fact that the equity hedge fund industry makes up such a small portion of the total equity market.

The primary implication of the negative parabolic shape of fund-level returns to scale is that many hedge funds are underfunded. This theoretical dilemma can be reformulated as a practical investment strategy that is long IRS funds and short DRS funds. Such a strategy yields positive returns on average and is robust to various performance measures and different identification methods. Importantly, the identification method which uses nonlinear functional form captures the scale effect more accurately and generates higher returns.

References

1. Agarwal, V., Daniel, N. D. & Naik, N. Y. Flows, performance, and managerial incentives in hedge funds (2004).
2. Agarwal, V., Green, T. C. & Ren, H. Alpha or beta in the eye of the beholder: What drives hedge fund flows? *Journal of Financial Economics* **127,** 417–434 (2018).
3. Aggarwal, R. K. & Jorion, P. Hidden survivorship in hedge fund returns. *Financial Analysts Journal* **66,** 69–74 (2010).
4. Ammann, M. & Moerth, P. Impact of fund size on hedge fund performance. *Journal of Asset Management* **6,** 219–238 (2005).
5. Angrist, J. D. & Pischke, J.-S. Instrumental variables in action: sometimes you get what you need. *Mostly harmless econometrics: an empiricist's companion,* 113–220 (2009).
6. Berk, J. B. & Green, R. C. Mutual fund flows and performance in rational markets. *Journal of political economy* **112,** 1269–1295 (2004).
7. Bollen, N. P. & Pool, V. K. Do hedge fund managers misreport returns? Evidence from the pooled distribution. *The Journal of Finance* **64,** 2257–2288 (2009).
8. Brown, S. J. & Goetzmann, W. N. *Hedge funds with style* tech. rep. (National Bureau of Economic Research, 2001).
9. Cao, C. & Velthuis, R. *Are Hedge Fund Capacity Constraints Binding? Evidence on Scale and Competition* tech. rep. (Working paper, Pennsylvania State University, 2017).
10. Chen, J., Hong, H., Huang, M. & Kubik, J. D. Does fund size erode mutual fund performance? The role of liquidity and organization. *American Economic Review* **94,** 1276–1302 (2004).
11. Chung, J.-W. & Kang, B. U. Prime broker-level comovement in hedge fund returns: information or contagion? *The Review of Financial Studies* **29,** 3321–3353 (2016).

12. Fama, E. F. & French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of financial economics* **33,** 3–56 (1993).

13. Ferreira, M. A., Keswani, A., Miguel, A. F. & Ramos, S. B. The determinants of mutual fund performance: A cross-country study. *Review of Finance* **17,** 483–525 (2013).

14. Fung, W. & Hsieh, D. A. Empirical characteristics of dynamic trading strategies: The case of hedge funds. *The review of financial studies* **10,** 275–302 (1997).

15. Fung, W. & Hsieh, D. A. Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases. *Journal of Financial and Quantitative analysis* **35,** 291–307 (2000).

16. Fung, W. & Hsieh, D. A. The risk in hedge fund strategies: Theory and evidence from trend followers. *The review of financial studies* **14,** 313–341 (2001).

17. Fung, W. & Hsieh, D. A. Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal* **60,** 65–80 (2004).

18. Fung, W., Hsieh, D. A., Naik, N. Y. & Ramadorai, T. Hedge funds: Performance, risk, and capital formation. *The Journal of Finance* **63,** 1777–1803 (2008).

19. Fung, W., Hsieh, D., Naik, N. & Teo, M. Growing the asset management franchise: evidence from hedge fund firms (2016).

20. Getmansky, M. The life cycle of hedge funds: Fund flows, size, competition, and performance. *The Quarterly Journal of Finance* **2,** 1250003 (2012).

21. Goetzmann, W. N., Ingersoll Jr, J. E. & Ross, S. A. High-water marks and hedge fund management contracts. *The Journal of Finance* **58,** 1685–1718 (2003).

22. Hjalmarsson, E. Predicting global stock returns. *Journal of Financial and Quantitative Analysis* **45,** 49–80 (2010).

23. Joenväärä, J., Kosowski, R. & Tolonen, P. Revisiting stylized facts about hedge funds. *Imperial College Business School* (2012).

24. Joenväärä, J., Kosowski, R. & Tolonen, P. The effect of investment constraints on hedge fund investor returns. *Journal of Financial and Quantitative Analysis,* 1–64 (2013).

25. Jorion, P. & Schwarz, C. Are hedge fund managers systematically misreporting? Or not? *Journal of Financial Economics* **111,** 311–327 (2014).

26. Jorion, P. & Schwarz, C. The strategic listing decisions of hedge funds. *Journal of Financial and Quantitative Analysis* **49,** 773–796 (2014).

27. Moon, H. R. & Phillips, P. C. Estimation of autoregressive roots near unity using panel data. *Econometric Theory* **16,** 927–997 (2000).

28. Naik, N. Y., Ramadorai, T. & Stromqvist, M. Capacity constraints and hedge fund strategy returns. *European Financial Management* **13,** 239–256 (2007).

29. Pástor, L., Stambaugh, R. F. & Taylor, L. A. Scale and skill in active management. *Journal of Financial Economics* **116,** 23–45 (2015).

30. Roberts, M. R. & Whited, T. M. in *Handbook of the Economics of Finance* 493–572 (Elsevier, 2013).

31. Sul, D., Phillips, P. C. & Choi, C.-Y. Prewhitening bias in HAC estimation. *Oxford Bulletin of Economics and Statistics* **67,** 517–546 (2005).

32. Teo, M. Does size matter in the hedge fund industry? *Available at SSRN 1331754* (2009).

33. Yan, X. S. Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis* **43,** 741–767 (2008).

34. Zhong, Z. Why does hedge fund alpha decrease over time? Evidence from individual hedge funds (2008).

35. Zhu, M. Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics* **130,** 114–134 (2018).

A Tables

Table 1: Hedge Fund Factor Loadings

This table gives the summary statistics for the FH factors and the CAPM factor. The eight factors used to calculate FH alpha are S&P 500 return minus risk free rate (MKT), Wilshire small-cap minus large-cap return (SZSPRD), yield spread of the U.S. 10-year Treasury bond over the three-month Treasury bill, adjusted for the duration of the U.S. 10-year Treasury bond (BNDMKT), change in the credit spread of Moody's BAA bond over 10-year Treasury bond adjusted for duration (CRDSPRD), excess returns on portfolios of lookback straddle options on currencies PTFS (PTFSFX), bonds (PTFSBD), and commodities PTFS (PTFSCOM), and Emerging Market Index total return (EMRG). The factor used to calculate CAPM alpha is excess market return (MKTRF), as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate. All values are given as percentages.

	Mean	Sd	Skewness	Kurtosis	1%	25%	50%	75%	99%
MKT	0.64	4.22	-0.68	4.23	-11.01	-1.66	1.01	3.45	9.31
SZSPRD	0.04	3.26	0.28	7.64	-7.03	-2.19	0.04	2.09	7.89
BNDMKT	-0.01	0.25	0.00	4.34	-0.61	-0.17	0.00	0.13	0.64
CRDSPRD	0.00	0.21	1.30	14.44	-0.65	-0.10	0.00	0.08	0.57
PTFSFX	-0.62	18.81	1.16	4.49	-29.74	-14.82	-4.34	8.86	66.01
PTFSBD	-2.27	14.62	1.34	5.60	-23.92	-12.86	-4.77	3.50	45.21
PTFSCOM	-0.37	14.23	1.08	4.60	-22.94	-9.82	-3.14	6.63	41.44
EMRG	0.69	6.51	-0.69	5.17	-17.49	-2.63	0.85	4.76	14.38
MKTRF	0.66	4.35	-0.73	4.23	-10.72	-1.95	1.18	3.44	9.54

Table 2: Descriptive Statistics

This table gives a summary of the main quantities used in the analysis. Net Return is given in the TASS database. From this measure, Gross Return is backed out. Common hedge fund risk factors are used to create two alpha measures, FH alpha and CAPM alpha. Fund Size is shown in millions of USD, adjusted for inflation to January 2000 dollars. Industry Size is shown using two measures. The first is in billions of USD, as the sum of all real fund sizes at time t. The second is sum of all nominal fund sizes at time t divided by the total US stock market capitalization at time t. All variables are winsorized at 1% and 99%.

	No. Funds/month	Mean	Sd	Skewness	Kurtosis	1%	25%	50%	75%	99%
Net Return	192,340	0.69	4.11	-0.05	5.50	-12.86	-0.98	0.66	2.40	13.73
Gross Return	186,524	2.02	4.40	0.09	5.42	-11.99	0.13	1.93	3.88	16.44
FH Alpha	142,336	1.68	2.58	0.16	4.54	-5.96	0.35	1.63	2.95	9.92
CAPM Alpha	142,336	1.64	3.41	0.21	5.06	-8.81	0.01	1.57	3.16	13.05
Fund Size (mil. USD)	152,844	120.65	246.45	4.44	26.23	0.46	12.26	38.44	111.52	1552.25
Industry Size (bil. USD)	192,340	80.60	41.42	0.42	2.06	11.63	46.58	72.03	119.37	164.94
Industry Size (%)	192,340	0.56	0.28	-0.13	1.49	0.12	0.27	0.60	0.84	0.95

Table 3: Comparison of Recursive Demeaning Methodologies

This table shows the regression results of gross return on real size, using linear functional form in Panel A and loglinear functional form in Panel B. The first model is the recursive demeaning procedure used previously in the literature (RD1) and the second model is the enhanced recursive demeaning procedure (RD2). The estimates for the coefficients of size are reported. The intercepts of the first stage of the 2SLS regression are given along with the R-squared values from the first stage and the F-statistics. Heteroskedasticity-robust standard errors double-clustered by category x month are reported in the parentheses. The robust standard errors are clustered by fund as well.

	RD1	RD2
Panel A: Dollar Fund AUM		
Estimated Scale Effect	-1.08e-09	-6.53e-09**
	(2.82e-09)	(2.56e-09)
1st Stage Intercept	0	-1.67e07***
	-	(2.42e06)
1st Stage R2 (%)	0.50	3.48
1st Stage F-statistic	3.47	24.78
Panel B: Logarithm of Dollar Fund AUM		
Estimated Scale Effect	0.760	-0.252
	(3.97)	(0.157)
1st Stage Intercept	0	-2.00***
	-	(0.151)
1st Stage R2 (%)	0.02	7.26
1st Stage F-statistic	0.78	90.71
N obs	147,153	147,153
N funds	2,099	2,099

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Test of Returns-to-Scale Effects based on Gross Returns

This table shows the regression results of gross return on log size, augmented with squared log size, and then augmented with industry size as measured in two different ways for robustness. In the column labeled (3.1) industry size is measured by total AUM of equity funds at a time t divided by total US stock market capitalization at time t. In the column labeled (3.2), industry size is measured by total AUM of equity funds at a time t. Heteroskedasticity-robust standard errors double-clustered by category x month are reported in the parentheses. The robust standard errors are clustered by fund as well.

	(1)	(2)	(3.1)	(3.2)
Log Size	-0.252	3.665**	4.450**	4.036**
	(0.157)	(1.634)	(1.859)	(1.784)
Squared Log Size		-0.116**	-0.142***	-0.128**
		(0.0468)	(0.0548)	(0.0522)
Industry Size			-8.554	-2.93e-12
			(56.95)	(3.42e-12)
N obs	147,153	147,153	147,153	147,153
N funds	2,099	2,099	2,099	2,099

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Test of Returns-to-Scale Effects based on Fung-Hsieh Alpha

This table shows the regression results of FH alpha on log size, augmented with squared log size, and then augmented with industry size as measured in two different ways for robustness. In the column labeled (3.1) industry size is measured by total AUM of equity funds at a time t divided by total US stock market capitalization at time t. In the column labeled (3.2), industry size is measured by total AUM of equity funds at a time t. Heteroskedasticity-robust standard errors double-clustered by category x month are reported in the parentheses. The robust standard errors are clustered by fund as well.

	(1)	(2)	(3.1)	(3.2)
Log Size	0.602***	6.161***	6.912***	6.577***
	(0.133)	(1.207)	(1.274)	(1.247)
Squared Log Size		-0.161***	-0.184***	-0.175***
		(0.0330)	(0.0350)	(0.0342)
Industry Size			-13.26	-5.55e-13
			(17.21)	(1.06e-12)
N obs	112,088	112,088	112,088	112,088
N funds	2,017	2,017	2,017	2,017

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Test of Returns-to-Scale Effects based on CAPM Alpha

This table shows the regression results of CAPM alpha on log size, augmented with squared log size, and then augmented with industry size as measured in two different ways for robustness. In the column labeled (3.1) industry size is measured by total AUM of equity funds at a time t divided by total US stock market capitalization at time t. In the column labeled (3.2), industry size is measured by total AUM of equity funds at a time t. Heteroskedasticity-robust standard errors double-clustered by category x month are reported in the parentheses. The robust standard errors are clustered by fund as well.

	(1)	(2)	(3.1)	(3.2)
Log Size	0.628***	8.546***	8.775***	9.000***
	(0.236)	(1.875)	(1.958)	(1.962)
Squared Log Size		-0.229***	-0.241***	-0.246***
		(0.0498)	(0.0526)	(0.0526)
Industry Size			37.16	4.09e-13
			(30.36)	(2.04e-12)
N obs	112,088	112,088	112,088	112,088
N funds	2,017	2,017	2,017	2,017

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Long/Short Portfolio of IRS and DRS Funds

This table shows the results of the long/short portfolio analysis of IRS and DRS funds. Panel A gives the results for the first identification method for IRS/DRS funds and Panel B gives the results for the second method. After using rolling regressions, funds are characterized as either IRS or DRS with a dummy variable. The dummy variable is then lagged so that for each period t, we have an indicator of whether the fund was experiencing IRS or DRS in period $t-1$. Then, for each month t we build a value-weighted portfolio of IRS funds and a value-weighted portfolio of DRS funds. We subtract the returns of the DRS portfolio from the returns of the IRS portfolio to create a long/short return measure that is long IRS funds and short DRS funds (LS = IRS - DRS). This measure is then regressed on a constant to measure average return and significance.

	FH Alpha	CAPM Alpha	Gross Return
Panel A: Method 1			
Mean Return	0.0421	0.104*	0.0427
	(0.0366)	(0.0549)	(0.0740)
IRS (%)	33.76	21.64	17.02
DRS (%)	66.24	78.36	82.98
N obs	97,507	97,507	132,966
N months	244	244	265
Panel B: Method 2			
Mean Return	0.0876**	0.145***	0.0685
	(0.0348)	(0.0521)	(0.0668)
IRS (%)	38.34	34.10	30.89
DRS (%)	61.66	65.90	69.11
N obs	92,199	92,199	125,098
N months	245	245	266

Standard deviations in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Figures

Figure 1: Negative parabolic shape of the size-performance relationship. Funds experience increasing returns to scale until they reach an optimal size, at which they begin experiencing decreasing returns to scale.