A Bottom-up Estimate of Decreasing Returns to Scale *

Jiacui Li and Min Zhu

August 28, 2026

ABSTRACT

In understanding active fund management, the explanatory power of the Berk and Green (2004) framework hinges on the magnitude of decreasing returns to scale (DRS) — the degree to which fund size erodes performance. We develop a bottom-up framework that decomposes fund alpha into economically interpretable sources of DRS. Applying the decomposition to U.S. stock mutual funds, we estimate that for an average fund, doubling fund size leads to an approximately 0.3% decline in annual returns. DRS is mainly driven by less nimble portfolio adjustments and higher transaction costs as funds grow. Magnitudes are heterogeneous, but even among funds with the highest DRS they remain substantially smaller than some prior estimates. Our findings shed light on the quantitative relevance of the Berk and Green mechanism.

JEL classification: G11, G23, J24
Keywords: Decreasing returns to scale; mutual fund; managerial skill

*Li (jiacui.li@eccles.utah.edu) is at David Eccles School of Business, University of Utah. Zhu (m.zhu@business.uq.edu.au) is at University of Queensland Business School. Zhu acknowledges support from the Australian Research Council Discovery Project (DP240100277). The authors are grateful for comments from John Adams, Azi Ben-Rephael, Aditya Chaudhry, Benjamin Golez, Matthew Gustafson, Da Huang, Matthew Jaremski, Zihan Lin, Xu Lu, Lubos Pastor, Kuntara Pukthuanthong, Jonathan Reuter, Andrea Rossi, Yang Shi, Sophie Shive, Yang Song, Yuehua Tang, Eric Zitzewitz , as well as seminar participants at BYU, Deakin, Northeastern, RMIT, UC Boulder, UNSW, Utah, Utah State University . The authors are particularly grateful for many insightful discussions with Konark Saxena in an earlier version.

1 Introduction

There are several striking stylized facts in the world of active mutual funds: active funds deliver low or no alpha on average, and realized alphas do not persist. The traditional view is that managers have limited skill and variations in realized alphas are mostly due to luck (Jensen, 1968; Malkiel, 1995; Carhart, 1997; Fama and French, 2010). Berk and Green (2004) (henceforth BG), in contrast, offer an alternative interpretation based on decreasing returns to scale (DRS) — the insight that funds deliver lower percentage alpha when their sizes grow. Under the BG interpretation, fund managers are skilled, and the aforementioned stylized facts are due to DRS interacting with performance-chasing flows (e.g. Sirri and Tufano, 1998; Chevalier and Ellison, 1997).

While BG's qualitative insights go through whenever DRS is present, their *quantitative* explanatory power hinges on the *magnitude* of DRS. Researchers have estimated DRS of active equity mutual funds using a variety of statistical methods and end up with a wide range of estimates. To put magnitudes into perspective, we express DRS as the decline in annual alpha when fund size doubles. High-end estimates such as Zhu (2018) and Van Binsbergen et al. (2025) are as large as 3%, implying that a fund with 1% alpha would fall to -2% upon size doubling. Low-end estimates such as Barras et al. (2022) are an order of magnitude smaller, around 0.2–0.4%. Studies that exploit exogenous size variation often cannot reject zero DRS (Phillips et al., 2018; Reuter and Zitzewitz, 2021). Thus, despite extensive work, there remains limited consensus on how large DRS is — and therefore on how much of the stylized facts BG can quantitatively explain.

This paper provides a complementary, bottom-up approach to quantify DRS. We use accounting identities to decompose fund alpha into economically interpretable components and quantify the DRS in each component. Adding up those DRS components, for a typical active U.S. equity fund, we estimate total DRS to be around 0.33%. We find that DRS exhibits heterogeneity, and funds with higher turnover and more illiquid holdings suffer larger

DRS effects. However, even those with high DRS only experience an effect in the vicinity of 0.5%. Applying our estimates, we find that while the BG mechanism certainly plays a role, it accounts for a limited share of the stylized facts for typical U.S. equity funds.

Our approach begins by decomposing the pre-fee alpha of a fund over quarter t into two components using an accounting identity:

$$\alpha_t^{\text{fund}} = \underbrace{\sum_j w_{j,t-1} \cdot \alpha_{j,t}}_{\text{holdings alpha}} + \underbrace{\left(\alpha_t^{\text{fund}} - \sum_j w_{j,t-1} \cdot \alpha_{j,t} \right)}_{\text{Kacperczyk et al. (2008) return gap}} \tag{1}$$

where the first component, holdings alpha, can be computed using the beginning-of-quarter portfolio weights $w_{j,t-1}$ and security-level alphas $\alpha_{j,t}$. It captures the alpha that could have been achieved if the fund fixed its positions over the quarter and incurred no execution costs. Quarterly holdings from 13F filings help us discipline this component. The second term, the return gap, which we compute following Kacperczyk et al. (2008), reflects unobserved intraquarter trading, transaction costs, as well as lending and financing fees. With this decomposition, we can quantify the DRS effects on each component in turn and add them up to get the total DRS effect.

We first quantify DRS in holdings alpha. Further accounting identities show that when a fund's size increases, its holdings alpha can decline for two reasons: (i) "dilution", as portfolio weights become less concentrated on the highest-alpha ideas (e.g. Pollet and Wilson, 2008; Pástor et al., 2020; Busse et al., 2021); and (ii) "alpha erosion", which refers to declines in security-level alphas due to tilting toward more liquid stocks where alpha opportunities are scarcer (e.g. Pástor et al., 2020) and due to larger fund positions creating bigger price impacts (e.g. Coval and Stafford, 2007; Lou, 2012; Davis et al., 2026). By examining how 13F holdings ($w_{j,t-1}$) change with fund size, we find that while funds do change their behavior

when size increases, the changes are modest.[1] Putting the pieces together, we estimate holdings-alpha DRS to be about 0.09% per size doubling for a typical fund.

We then turn to the return gap component. Regressing it on lagged fund size with fund fixed effects, we find an average DRS of about 24 bp — almost three times the holdings-alpha DRS effect, and thus the main source of total DRS. To further break down the source of DRS in the return gap, we note that the return gap component of return primarily arises from intraquarter trading and trading costs (e.g. Perold and Salomon, 1991; Edelen et al., 2013; Pástor et al., 2017), and the other parts are minor. Both parts have their own DRS effects: as size increases, funds become less nimble at intraquarter trading and incur larger transaction costs.

Putting together the two components — holdings alpha and return gap — we estimate that the total DRS effect is about $0.09\% + 0.24\% = 0.33\%$ per size doubling for a typical fund. This is broadly in line with the low-end statistical estimates and much lower than the high-end estimates.

We also find DRS to be heterogeneous across funds, a theme emphasized in the prior literature (e.g. Chen et al., 2004; Yan, 2008; Barras et al., 2022; Evans et al., 2023; Van Binsbergen et al., 2025). Prior papers document the pattern statistically, and we complement them by deriving it from economic mechanisms. In our decomposition, both holdings-alpha and return-gap DRS are larger for funds with higher turnover or less liquid holdings, because those characteristics amplify price impact, trading costs, and related frictions. For funds with high DRS, we estimate the effect to be about 0.5% of annual alpha per size doubling.

Finally, we use our estimates to shed light on the quantitative explanatory power of the BG framework. In particular, BG argue that fund performance has limited persistence because

[1]For instance, to evaluate the liquidity tilt channel, we use each stock's NYSE size decile as a proxy for liquidity, and find that the portfolio value-weighted decile of holdings rises by only about 0.05 when fund size doubles (e.g., from decile 8 to 8.05).

flows chase winners (e.g. Chevalier and Ellison, 1997; Sirri and Tufano, 1998) and DRS erodes their edge. With our estimated DRS effects, the size adjustment mechanism biases the annual AR(1) coefficient of fund alpha by about 0.007. Relatedly, the share of cross-sectional alpha variance attributable to "skill" — the predictable component of alpha — is also largely unaffected by removing the effects of fund size adjustment. Thus, while the BG mechanism certainly operates, it accounts for a modest share of the stylized facts among U.S. active equity funds.

The main contribution of this paper is developing a bottom-up method to estimate the magnitude of DRS. With the help of holdings data, we find that DRS effects, while clearly present, are modest in U.S. equity mutual funds, which sheds light on the debate between the traditional and BG perspectives for understanding funds. Beyond a point estimate, the approach yields a portable decomposition that maps DRS into observable portfolio and trading frictions, clarifying which channels matter and how effects vary across funds with different characteristics.

We emphasize two caveats in interpreting our results. First, we view our approach as *complementary* to statistical methods. Statistical approaches have their own benefits, such as lower data requirements, and some estimate granular fund-by-fund DRS effects; ours instead links DRS to concrete economic mechanisms and disciplines estimates with holdings data. Second, while we find DRS effects to be modest in U.S. equities, which are highly liquid, it is possible — and perhaps even likely — that DRS is larger in less liquid markets such as venture capital, private equity, and hedge funds. If so, BG would be of more quantitative importance for describing those markets.

The rest of the paper is organized as follows. Section 2 presents our framework that decomposes fund alpha into holdings alpha and the return gap component. Section 3 describes the data. Section 4 quantifies DRS in holdings alpha, and Section 5 quantifies DRS in the return gap. Section 6 discusses implications of these magnitudes for the BG interpretation

of mutual fund facts. Section 7 concludes.

1.1 Related literature

The debate about how to interpret the low active mutual fund alpha and persistence traces back to the seminal work of Berk and Green (2004) (BG). Before BG, the conventional wisdom holds that, because realized fund alphas are low and non-persistent, fund managers have limited skill and realized performance is mostly due to luck (Jensen, 1968; Malkiel, 1995; Carhart, 1997; Fama and French, 2010). BG, in contrast, show that when funds face decreasing returns to scale (DRS) and when competitive fund flows rationally chase performance (e.g. Sirri and Tufano, 1998; Chevalier and Ellison, 1997), fund flows adjust to the point of zeroing out post-fee alphas, which can explain all of the aforementioned facts simultaneously.

Since the seminal work of BG, many papers have attempted to estimate DRS using fund size and performance data through different statistical methods. The utilized methodologies include ordinary least squares (Chen et al., 2004; Yan, 2008; Ferreira et al., 2013), fixed effects and recursive demeaning (Pástor et al., 2015; Zhu, 2018; Huang et al., 2023), as well as random effects models (Harvey and Liu, 2021; Barras et al., 2022). Some researchers find instruments to isolate exogenous variation in fund sizes to estimate DRS (Phillips et al., 2018; Reuter and Zitzewitz, 2021). As summarized in Appendix A, these papers end up with a wide range of DRS estimates. Some papers argue that existing DRS estimates suffer from outliers (Adams et al., 2018a,b; Pástor et al., 2022). In addition to statistical estimates of DRS, some other papers examine fund portfolio behavior that shed light on the existence of DRS (Pollet and Wilson, 2008; Pástor et al., 2020; Busse et al., 2021).

To be closely aligned with the original BG paper, our paper focuses on *fund-level* DRS effects: how the performance of a fund varies with its own size. There is also a literature that examines whether fund performance varies with the size of the entire mutual fund

industry (Pástor and Stambaugh, 2012; Pástor et al., 2015). Our paper does not focus on industry-level DRS, and Appendix F.1 provides reasoning and evidence that industry-level effects do not impact our main conclusions.

2 A Bottom-up Framework for DRS

This section presents a framework to decompose DRS into economically interpretable components.

2.1 A decomposition of DRS

DRS is defined as how fund alpha changes with fund size. We start with an approximate accounting identity for fund alpha (derived in Appendix C.1). Note that the identity holds regardless of the right model for measuring alpha.

Consider the alpha of a fund over a time period [0,1], such as a quarter, which can be further decomposed into K subperiods (e.g. days). Let $j = 1, ..., N$ denote securities, and let α_j denote the alpha of security j over the whole period and $\alpha_{j,k+1}$ denote its alpha over subperiod $[k, k+1)$. Let $w_{j,k}$ denote the portfolio weight at the start of subperiod $k+1$, and let $\text{TCost}_{j,k+1}$ denote the trading costs incurred by security j in period $[k, k+1)$. Then, fund alpha is given by an identity:

$$\alpha_t^{\text{fund}} = \sum_{k=0}^{K-1} \underbrace{\sum_j w_{j,k} \cdot (\alpha_{j,k+1} - \text{TCost}_{j,k+1})}_{\text{Alpha over period } k} + \text{Residual} \tag{2}$$

Thus, if we observe all intra-period holdings, we know fund alpha up to two adjustments. The first is transaction costs ($\text{TCost}_{j,k+1}$), which include commissions, bid-ask spreads, and temporary price impact. The second is a residual that captures net lending fees, financing costs, and compounding approximation errors. As shown in Appendix B.1, active mutual

funds engage in little securities lending and rarely use shorting or leverage, making lending fees and financing costs small. The compounding error arises because simple returns do not aggregate additively over time; Appendix B.2 confirms this has a limited effect on DRS.

To separate fund alpha into two components that we can measure and study separately, we rearrange (2) as:

$$\alpha_t^{\text{fund}} = \underbrace{\sum_j w_{j,0}\alpha_j}_{\text{(1) holdings alpha}} + \underbrace{\overbrace{\sum_j \sum_{k=0}^{K-1} (w_{j,k} - w_{j,0}) \cdot \alpha_{j,k+1}}^{\text{weight deviation}} - \overbrace{\sum_j \sum_{k=0}^{K-1} w_{j,k}\text{TCost}_{j,k+1}}^{\text{trading costs}} + \text{Residual}}_{\text{(2) Kacperczyk et al. (2008) return gap}} \quad (3)$$

where $w_{j,0}$ is the portfolio weight at the start of the period, and $\alpha_j \approx \sum_{k=0}^{K-1} \alpha_{j,k+1}$ is the full-period alpha of security j.

We compute both components quarterly. Holdings alpha is calculated using quarterly 13F disclosures and security alphas from standard asset pricing models. The second component, the return gap (Kacperczyk et al., 2008), is the difference between the fund's actual gross alpha (α_t^{fund}) from CRSP/Morningstar and its holdings alpha. As in Kacperczyk et al. (2008), holdings alpha explains most of the fund performance variance.

Per this accounting identity in (3), estimating DRS can be separated into understanding how these two components vary with fund size. Sections 4 and 5 analyze them in turn, estimate the DRS that arise from each component, and compare the estimates to relevant studies in the literature.

2.2 Why use a bottom-up approach?

The literature has used several statistical methods to estimate the relationship between fund size and performance, as discussed in Section 1.1. This section explains why we introduce our new approach and discusses its benefits and drawbacks.

First, many statistical methodologies have been proposed to estimate DRS from the relationship of fund size and performance, but they yield a wide range of estimates even when applied on largely the same sample of active U.S. equity mutual funds. Appendix A surveys those estimates which we summarize here. For interpretation, we convert estimates into a common unit: the decline in annual alpha of a typical active fund if its size doubles (e.g. from 300 million to 600 million USD). The estimates vary wildly. The high-end estimates can be above three percentage points — that is, if a fund had 1% annual alpha initially, it will decline to lower than -2%.[2] The low-end estimates imply that alpha only declines by around 0.2%. Further, studies that utilize exogenous variation in fund sizes cannot reject the null hypothesis of zero DRS. We believe that this large range of estimates calls for an alternative approach as a supplemental check.

Second, our approach provides economic interpretations to the different components of DRS. Specifically, the decomposition links DRS to concrete portfolio and trading frictions — most importantly, less nimble intraquarter trading and higher transaction costs as funds grow. That mapping also clarifies heterogeneity: funds with higher turnover or less liquid holdings should experience more pronounced DRS. Having interpretable components further lets us discipline our estimates with evidence on fund portfolio behavior from related studies.

Third, our method makes use of holdings and turnover data to help discipline estimates. Existing approaches have primarily relied on fund size and performance data to estimate DRS. Of course, DRS does not happen in a vacuum — it arises as funds adjust their holdings and trading behavior in response to size changes. Using holdings data also allows us to reconcile our bottom-up findings with documented patterns of portfolio behavior (e.g. Pollet and Wilson, 2008).

[2]It is worth pointing out that this is a very large estimate, especially given those studies are often based on large size variation. Over the life of a fund, its size can easily vary by a factor of 10, and such a DRS estimate implies that its annual alpha varies in a range of $3\% \times \log_2(10) \approx 10\%$, which is a very large range, given that the typical fund has an annual pre-fee alpha of about 0.5 to 1% (Berk and Van Binsbergen, 2015).

Finally, we emphasize that we see the bottom-up approach as a complementary check for, rather than a replacement for, statistical approaches. The bottom-up approach also has its own drawbacks. First, by using holdings data, our approach has higher data requirements, while the statistical methods usually only require fund size and performance data. Second, the bottom-up approach also requires making functional form assumptions about the price impact of trading to arrive at the final estimates. As such, it will be more accurate to think of the bottom-up estimates as approximate rather than exact.

3 Data

We obtain monthly U.S. mutual fund flow, return, and benchmark data from CRSP from 1991 to 2022. We restrict attention to U.S. domiciled mutual funds that focus on domestic equity and are actively managed. We merge the data with Morningstar to obtain styles (e.g. large cap-growth), and use that to identify benchmarks and restrict the sample to remove undiversified sector funds. For each of the 3×3 size-value styles, Morningstar specifies a benchmark index which we use to compute benchmark-adjusted returns.

Much of our analysis relies on holdings. For this purpose, we merge the data with quarterly 13F fund holdings from Thomson Reuters using MFLINKS from Wermers (2000). We then use the code from Kacperczyk et al. (2008) to compute the return gap. To ensure the quality of matches, we focus on funds for which we find at least 30 portfolio holdings and that have a CRSP-reported stock fraction of holdings between 80% and 120%. To enable calculating fund alphas using factor model regressions, we also require funds to have at least 24 months of data. Our main conclusions are robust to more or less stringent data filters.

Table 1 presents summary statistics. After requiring holdings from 13F and styles from Morningstar, we end up with 2,305 unique funds and 281,700 fund-month observations. The median fund size is 326 million USD and the variable is right-skewed. Average annual

turnover is 76% and average annual expense ratio is 1.13%. Because we are interested in understanding the skill of managers rather than the returns obtained by fund investors, we focus on gross-of-fee returns and alpha throughout. We compute alphas by either using various time-series regressions of factor models — CAPM, FF3, etc — or by subtracting the Morningstar benchmark return for the specific 3×3 style that the fund belongs to. As expected, benchmark- or factor-adjusted returns are rather close to zero on average. Depending on the model used for adjustment, the monthly standard deviation of alpha is in a range of 1.56% to 2.34%, which annualizes to 5.40% to 8.11

When decomposing fund returns into holdings return and return gap (equation (3)), consistent with Kacperczyk et al. (2008), we find that holdings return can explain most of the variation in fund returns. Return gap has a mean of -0.01% per month, which is close to zero, and a standard deviation of 0.75% which annualizes to 2.6%. In terms of return variance explained, this amounts to $\frac{0.75^2}{5.38^2} \approx 2\%$ of overall fund return variation.

			Percentiles						
	Mean	StDev	1%	5%	25%	50%	75%	95%	99%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Assets ($m)	1,652	5,822	13	23	97	326	1,155	6,806	22,036
Annual turnover	0.76	0.68	0.03	0.11	0.33	0.59	0.98	1.93	3.22
Expense ratio (%)	1.13	0.39	0.19	0.51	0.90	1.09	1.33	1.83	2.22
Gross return (%)	0.79	5.38	-15.05	-8.63	-2.00	1.26	3.97	8.60	13.46
Benchmark return (%)	0.75	5.30	-15.15	-8.59	-1.93	1.31	3.87	8.54	13.21
Benchmark adjusted return (%)	0.04	1.81	-4.88	-2.60	-0.79	0.02	0.84	2.76	5.33
CAPM adjusted return (%)	0.03	2.34	-6.33	-3.52	-1.05	0.01	1.08	3.65	6.62
FF3 adjusted return (%)	0.02	1.70	-4.62	-2.59	-0.83	0.02	0.86	2.61	4.78
FFC adjusted return (%)	0.01	1.63	-4.44	-2.53	-0.82	0.02	0.84	2.54	4.54
FF5 adjusted return (%)	0.02	1.63	-4.36	-2.47	-0.82	0.00	0.82	2.53	4.67
FF6 adjusted return (%)	0.01	1.56	-4.15	-2.39	-0.80	0.00	0.80	2.45	4.42
Holdings return (%)	0.80	5.56	-15.69	-8.84	-2.08	1.29	4.07	8.85	13.85
Return gap (%)	-0.01	0.75	-2.09	-0.97	-0.27	-0.01	0.25	0.98	2.17

Table 1. Summary statistics

This table summarizes our monthly sample of diversified active U.S. equity funds over 1991 to 2022, containing a total of 2,305 funds and 281,700 fund-month observations. The first two columns report the mean and standard deviation of variables, while the remaining columns report percentile distributions. The gross and adjusted returns, as well as return gap, are of monthly frequency while the turnover and expense ratios are of annual frequency. Return gap calculation follows Kacperczyk et al. (2008), where "holdings return" is the return that would have been obtained if the fund had held the same positions throughout the period, and return gap is the difference between gross fund return and holdings return. CAPM, FF3, FFC, FF5, and FF6 stand for the Capital Asset Pricing Model, Fama-French three-factor model, Fama-French-Carhart four-factor model, Fama-French five-factor model, and Fama-French six-factor model, respectively.

4 DRS in Holdings Alpha

This section quantifies DRS effects on holdings alpha.

4.1 Decomposing DRS in Holdings Alpha

We use another accounting identity to further decompose the DRS in holdings alpha into two interpretable parts. Consider a fund choosing positions. Intuitively, it will rank its security picks ("ideas") in descending order $r = 1, \ldots, N$, and it will tend to allocate more weights to its highest-conviction ideas, balancing alpha and diversification. To capture this, we express

holdings alpha as:

$$\text{HA} = \sum_{r=1}^{N} w_r \cdot \alpha_r$$

where $r = 1$ means the best idea, $r = 2$ means the next-best idea, etc., and recall N is the total number of stocks. Relative to the security index j in Section 2, here r reindexes the same holdings by conviction rank rather than by security identity. To study the effect of DRS, note that if fund size increases, the change in holdings alpha is given by:

$$\begin{aligned} \widetilde{\text{HA}} - \text{HA} &= \sum_r \tilde{w}_r \tilde{\alpha}_r - \sum_r w_r \alpha_r \\ &= \underbrace{\sum_r (\tilde{w}_r - w_r)\alpha_r}_{\text{dilution}} + \underbrace{\sum_r \tilde{w}_r (\tilde{\alpha}_r - \alpha_r)}_{\text{alpha erosion}} \end{aligned} \tag{4}$$

where $\tilde{w}_r$ and $\tilde{\alpha}_r$ are the portfolio weights and security alphas after the change, respectively.

The first term in (4) measures "dilution": as funds become larger, they may choose to reduce their positions in their highest-conviction ideas and spread their portfolios across more stocks (e.g. Pollet and Wilson, 2008). The second term measures "alpha erosion": as funds become larger, even if they do not reduce their portfolio concentration, the alphas of their holdings may decrease due to larger price impact or having to switch to choosing more liquid securities which often exhibit lower alphas (e.g. Pástor et al., 2020; Busse et al., 2021).

4.2 Dilution effect

To estimate the dilution term in (4), we need a model of $\{\alpha_r\}_{r=1}^{N}$: how security-level alpha declines with conviction. For this, we follow Shumway et al. (2009) and Anton et al. (2021) who use observed portfolio holdings to reverse engineer the conviction of fund managers on each "idea". Anton et al. (2021) present evidence that the top ideas exhibit high alphas

which decline sharply with conviction.

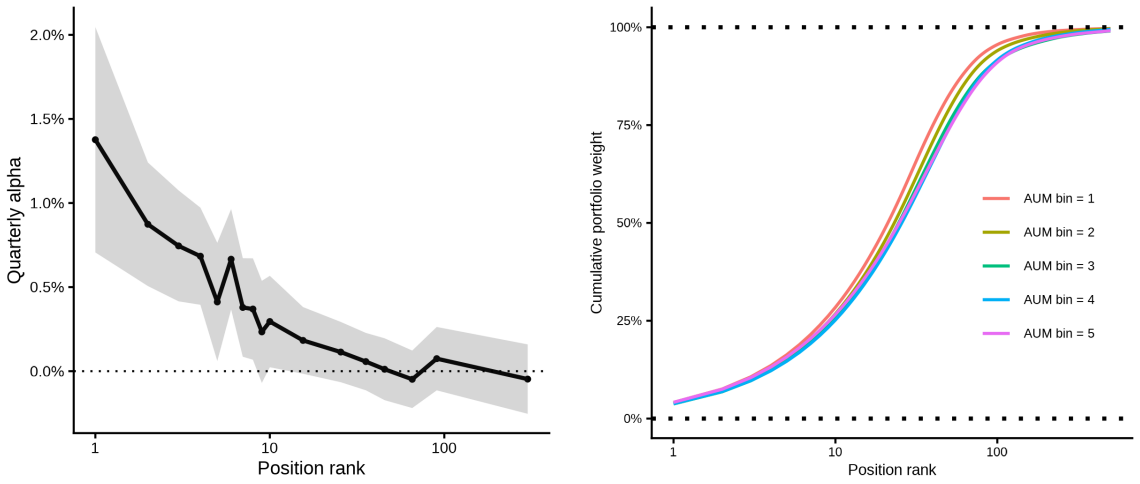
Alpha declines with conviction. As a baseline specification, we rank each fund's positions by the active holding relative to benchmark, $w_{i,t} - w^b_{i,t}$, where benchmark weight $w^b_{i,t}$ is proxied by the value-weighted portfolio of funds in the same Morningstar style benchmark.

Panel (a) of Figure 1 plots the average quarterly FF6 alpha of stocks by conviction ranking. Specifically, for each fund-quarter, we rank all stock holdings in descending order of conviction $r = 1, \ldots, N_i$ based on their active weights. We group these ranked positions into position-rank bin portfolios (e.g., $r = 1, 2, \ldots, 10$, and ranges thereafter). The plot reports the average FF6 alpha of those portfolios along with 95% confidence intervals.

The results are strongly consistent with Anton et al. (2021). The top-ranked idea of each fund creates around 1.38% of alpha by quarter, which declines steadily with the conviction number. The alpha ceases to be statistically significant when it comes to the 30th idea. We also tabulate the results in columns (1) and (2) in Table 2. The results show a rather monotonic pattern by which alpha declines, suggesting that this method of ranking positions is useful for identifying conviction. In Appendix D.1, we show that the general conclusion is robust to the factor model used to measure alpha, the exact way to identify conviction, and estimating the curve separately by fund-size quintile.

Funds diversify slightly when they become larger. To visualize how larger funds hold less concentrated portfolios, Panel (b) of Figure 1 plots the cumulative sum of portfolio weights by position rank. We sort funds in each quarter into 5 bins by assets under management (AUM) and plot them separately. For instance, funds in the smallest bin put 30.8%, 85.5%, and 96.1% of their portfolios in their top 10, 50, and 100 positions, respectively. Those numbers are also reported in column (3) of Panel A of Table 2.

As expected, larger funds indeed hold less concentrated portfolios, as the lines are stacked



(a) (b)

Figure 1. Alphas and positions by rank

For each fund-quarter, we rank its portfolio stocks in descending order of conviction, where conviction is measured by the deviation of the active holding from benchmark. Benchmark weight is proxied by the value-weighted portfolio of funds in the same Morningstar style. Panel (a) plots the average FF6 alpha of those stocks by conviction ranking with 95% confidence intervals. Panel (b) sorts funds into five bins by size (assets under management) and plots the cumulative sum of portfolio weights by position rank.

monotonically by fund size. However, the difference is not dramatic. As reported in Panel B of Table 2, funds in the top bin have median (mean) number of holdings of 85 (129) compared with the 58 (80) holdings for funds in the smallest bin, even though funds in the top bin are on average more than 100 times larger in AUM. This finding is consistent with Pollet and Wilson (2008) who find that while funds do appear to diversify when size increases, they do so rather slowly.

The combined dilution effect We now estimate the implied dilution effect from funds reducing portfolio concentration. We take the point estimate of alpha as a function of conviction ($\{\hat{\alpha}_r\}_{r=1}^{N}$) as given. For funds in each size bin b, we compute the "hypothetical alpha" as:

$$\widehat{HA}_b = \sum_r \hat{w}_{b,r} \cdot \hat{\alpha}_r$$

where $\hat{w}_{b,r}$ is the average portfolio weight of the r-th position in the b-th size bin. By examining how this varies by fund size bin b, we isolate the effect that only arises from dilution.

The last row of Table 2 reports them for each bin. From bin 1 to bin 5, even though fund size increased by a factor of $7,410/45 \approx 165$, the hypothetical alpha only decreased by 3 basis points per quarter which annualizes to 12 basis points. If we convert this to our standard metric, when fund size doubles, the effect on annual alpha is $0.12\%/\log_2(165) \approx 0.02\%$.

It is worth noting that this analysis compares *across* funds. Consistent with Pollet and Wilson (2008), within-fund comparisons also suggest that funds diversify only modestly as they grow. In Appendix D.2, we quantify the dilution-induced DRS using *within-fund* size variation in a panel regression of hypothetical alpha on lagged $\log_2$ AUM (with fund and time fixed effects, and standard errors clustered by fund and time) and find an even smaller effect of about 0.5 basis points per doubling. For the remainder of the paper, we take about 1 bp per doubling as a round baseline for the dilution channel.

4.3 Alpha erosion

We now turn to the second term, "alpha erosion," in equation (4). When funds grow larger, there are two potential reasons why their security-level alphas would decline. The first is price impact: as the fund becomes larger, it holds a larger fraction of available shares and pushes price up, which reduces subsequent alpha. The second is liquidity tilt: as funds become larger, they reduce exposure to small, illiquid stocks and migrate towards finding investment ideas in large, liquid stocks (Pástor et al., 2020; Busse et al., 2021). However, it is harder to find alpha opportunities in liquid stocks. This section seeks to quantify both and finds that the former is larger than the latter, but even the former is relatively small.

Panel A: Alpha and portfolio weight by position rank							
Position rank	Alpha (%)	SE	Cumulative port weights by fund AUM				
			Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1.38***	(0.34)	0.046	0.043	0.043	0.041	0.044
2	0.87***	(0.19)	0.083	0.079	0.077	0.075	0.079
3	0.75***	(0.17)	0.117	0.111	0.108	0.105	0.110
4	0.68***	(0.15)	0.148	0.140	0.137	0.132	0.138
5	0.41**	(0.18)	0.178	0.168	0.164	0.158	0.165
6	0.67***	(0.15)	0.206	0.194	0.189	0.182	0.190
7	0.38**	(0.15)	0.233	0.219	0.214	0.205	0.213
8	0.37**	(0.15)	0.259	0.244	0.237	0.227	0.236
9	0.23	(0.15)	0.284	0.267	0.259	0.249	0.258
10	0.30**	(0.14)	0.308	0.290	0.281	0.269	0.278
11-20	0.18*	(0.10)	0.517	0.486	0.468	0.445	0.452
21-30	0.11	(0.09)	0.679	0.639	0.612	0.582	0.583
31-40	0.06	(0.09)	0.788	0.750	0.719	0.687	0.682
41-50	0.01	(0.09)	0.855	0.823	0.793	0.767	0.756
51-80	−0.05	(0.09)	0.940	0.927	0.903	0.894	0.880
81-100	0.07	(0.10)	0.961	0.951	0.930	0.926	0.918
101-500	−0.05	(0.11)	0.999	0.997	0.992	0.995	0.992
501-3000	−0.08	(0.16)	1.000	1.000	1.000	1.000	1.000
Panel B: Hypothetical alpha by fund size							
			Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
AUM (m USD)			45	157	440	1,212	7,410
Num positions (median)			58	65	71	75	85
Num positions (mean)			80	97	120	112	129
Implied fund alpha ($\sum_j w_r \alpha_r$, %)			0.27	0.25	0.24	0.23	0.24

Table 2. Alpha and portfolio weights by position ranks

In Panel A, as described in Section 4.2, for each fund-quarter, we sort all its stock positions by descending conviction. Thus, the first row represents the highest-conviction position, the second row represents the second highest-conviction position, and so on. Column (1) reports quarterly FF6 factor alpha of those positions and column (2) reports the standard error. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01. In columns (3) through (7), we sort funds into five bins by AUM and report the cumulative portfolio weights by ranked stock position. In Panel B, the first row reports the average assets under management (AUM) of funds by bin; the next two rows report the median and mean number of holdings; and the last row reports the quarterly "hypothetical alpha" of these funds, defined as $\sum_{\text{rank } r} w(r) \cdot \alpha(r)$, where weights $w(r)$ and alphas $\alpha(r)$ are both taken from the estimates in Panel A.

Price impact We now quantify the price impact-induced effect on fund return when a fund's size doubles. This requires first modeling the price impact at the stock level and then aggregating it to the fund level.

For any stock j held by a fund, we model the log price of the stock as:

$$p_{j,t}(A) = p_{j,t}^{(0)} + \underbrace{f_j(A_t w_{j,t})}_{\text{price impact}} \tag{5}$$

where $p_{j,t}^{(0)}$ is the counterfactual stock price without the presence of the fund, A_t is the fund's AUM, $w_{j,t}$ is the fraction of the fund's portfolio allocated to the stock, and $f_j(\cdot)$ is the price impact function specific to this stock. For $f_j(\cdot)$, we use the estimates from Frazzini et al. (2018), who measure market impact from millions of live parent-order executions at a large asset manager. Live-trade calibrations of this kind are the standard industry approach to price impact (e.g., ITG); the resulting f_j is a concave function of trade size as a fraction of daily volume. Appendix D.3 describes the price impact estimates in more detail and also shows that our results are robust to using the price impact estimates from another study (Bershova and Rakhlin, 2013).

To translate this impact on stock *price* into the effect on stock *returns*, we plug in (5) into the Campbell-Shiller approximate identity:

$$\begin{aligned} r_{j,t+1} &\approx \kappa + \rho \cdot p_{j,t+1} + (1-\rho) d_{j,t+1} - p_{j,t} \\ &= \underbrace{\kappa + (1-\rho) d_{j,t+1} + \rho \cdot p_{j,t+1}^{(0)} - p_{j,t}^{(0)}}_{=r_{j,t+1}^{(0)}} - \underbrace{[f_j(A_t w_{j,t}) - \rho \cdot f_j(A_{t+1} w_{j,t+1})]}_{\text{return drag}} \end{aligned} \tag{6}$$

where κ is a constant term and $\rho \approx 0.96$ is the annual discount rate. $r_{j,t+1}^{(0)}$ is the counterfactual return when the fund has no impact. The second term is the return drag — the origin of price impact-induced DRS — due to price impact. Because we observe funds' AUM and portfolio weights, we can estimate this return drag using sample averages.

Before proceeding to estimates, we provide intuition for what determines the magnitude of the return drag. Obviously, the return drag is increasing in magnitude of fund holdings

$(A_t w_{j,t})$. The dependence on fund size A_t is what gives rise to DRS: larger funds have larger price impact. While not explicitly shown in the equation above, funds that operate in less liquid stocks will have larger $f_j(\cdot)$ functions and thus also face larger return drags.

The more interesting effect comes from the term $\rho \cdot f_j(A_{t+1} w_{j,t+1})$: it implies that the return drag is smaller if the fund's position is more persistent. Consider a fund whose position pushed up the price of a stock at time t. If it continues to hold the same position at time $t+1$, then the stock's price will continue to be elevated, and the negative effect on return is only a minor fraction $(1 - \rho \approx 0.04)$ of the current price impact. In contrast, if the fund entirely exits the position subsequently $(A_{t+1} w_{j,t+1} = 0)$, then the return drag is equal to the entirety of the current price impact. This implies that return drag is smaller for funds with lower turnover and thus higher position persistence. This mechanism is graphically illustrated in Figure 2.

We now aggregate the stock-level return drag in Equation (6) to the fund level:

$$\begin{aligned} \text{drag}(A_t) &\equiv -\sum_j w_{j,t} \cdot (r_{j,t+1} - r^{(0)}_{j,t+1}) \\ &= \sum_j w_{j,t} \cdot [f_j(A_t w_{j,t}) - \rho \cdot f_j(A_{t+1} w_{j,t+1})] \end{aligned} \tag{7}$$

We compute this quantity for each fund-quarter.

Finally, to infer the impact of fund size doubling, we compute $\text{drag}(1.5A) - \text{drag}(0.5A)$, which represents how much larger the return drag is when the fund doubles its size.[3] Table 3 reports the results by Morningstar size style and within-period AUM decile. As an example, the top-left cell is small-cap funds in the lowest AUM decile: on average 17 such funds each period, with mean AUM of $25 million. If a fund in this cell doubles its size around its

[3]Because the price impact is a concave function, the result does depend on the starting point. Therefore we choose to vary fund size around the current level. We keep the realized holdings path: the same 13F portfolio weights at t and the same subsequent positions at $t+1$, with exits mapped to zero subsequent impact; only dollar position size scales with A.

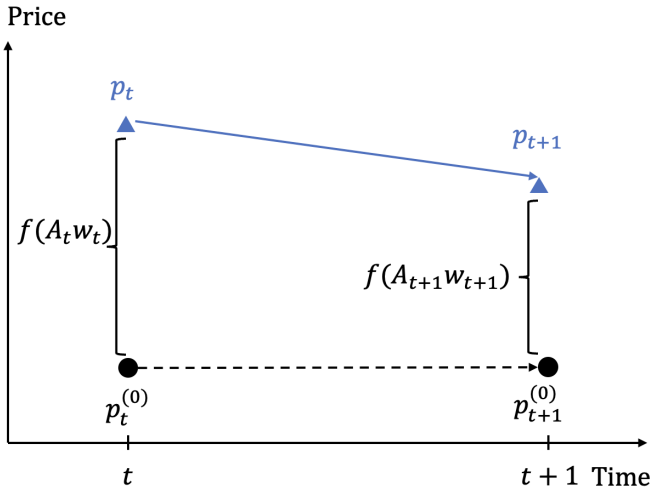


Figure 2. Translating price impact into the effect on stock returns

This figure illustrates how price impact translates into return drag under the stock price model of (5): $p_{j,t} = p_{j,t}^{(0)} + f_j(A_t w_{j,t})$, where $p_{j,t}^{(0)}$ is the counterfactual price when the fund has zero AUM (black dots) and $f_j(A_t w_{j,t})$ is the price impact due to a fund with AUM A_t allocating $w_{j,t}$ of its portfolio to stock j, and $p_{j,t}$ is the actual price (blue triangles). The vertical gaps are the price-impact levels $f_j(A_t w_{j,t})$ and $f_j(A_{t+1} w_{j,t+1})$. The black dashed line is the hypothetical price path without price impact. The solid blue line is the actual price change from t to $t+1$. Return drag can thus be visualized as the difference in slopes between the two lines. Note that return drag increases in the magnitude of current price impact ($f_j(A_t w_{j,t})$) and decreases in its persistence.

current AUM, annual return drag rises by 0.07%.

Across fund-quarters, the mean drag of size doubling is about 6 basis points per year. There is substantial heterogeneity by style and size. The smallest large-cap funds experience a return drag of only one basis point. Small-cap funds in the upper AUM bins — only a handful of them in an average period — experience a return drag of around 0.11–0.14% per year.

Liquidity tilt Another channel for alpha erosion is liquidity tilt. As funds become larger, they increase exposure to large, liquid stocks where alpha opportunities are harder to find.

We quantify the DRS from this channel in Appendix D.4 and summarize the results here. The contribution is small because the tilt itself is small. Concretely, we measure each fund's

	AUM bin									
Fund style	1	2	3	4	5	6	7	8	9	10
	Num of funds									
Small cap	17	16	16	16	17	17	15	11	8	4
Mid cap	14	12	13	13	13	12	13	15	13	10
Large cap	24	27	25	26	25	26	27	28	32	40
	AUM ($m)									
Small cap	25	56	102	173	273	441	707	1,156	2,024	5,221
Mid cap	24	58	102	170	274	439	699	1,151	2,042	5,852
Large cap	25	56	102	170	274	440	700	1,136	2,109	10,155
	Return drag (%)									
Small cap	0.07	0.09	0.09	0.10	0.11	0.12	0.12	0.14	0.14	0.11
Mid cap	0.03	0.04	0.05	0.05	0.07	0.09	0.09	0.10	0.11	0.12
Large cap	0.01	0.02	0.02	0.03	0.03	0.03	0.04	0.05	0.06	0.07

Table 3. Annual alpha erosion from price impact by fund style and size

This table reports the mean annual return drag from price impact when fund size doubles, by Morningstar size style and within-period AUM decile. Doubling is $\text{drag}(1.5A) - \text{drag}(0.5A)$ in (7), using the U.S. calibration of Frazzini et al. (2018), evaluated on each fund-quarter's realized 13F holdings path (portfolio weights held fixed; exits receive zero subsequent impact). The first two panels show the average number of funds and average AUM in each cell. The last panel is the mean annual return drag in percent.

portfolio value-weighted NYSE size decile of its stock holdings and find that doubling fund size raises this measure by only about 0.05 on average. That is, if a fund on average holds stocks at the 5th size decile, doubling its size will only raise the average size decile to 5.05. We further use the variation of characteristics-sorted long–short factor returns by stock size to gauge how achievable returns decline with stock size. Combining these two facts, the implied DRS from liquidity tilt is only about 1–2 bp per year when fund size doubles.

4.4 Putting estimates together: holdings alpha DRS

We now put together the estimates from the previous subsections to estimate the total DRS from holdings alpha. For a typical fund doubling its size, the impact on annual return is:

$$\text{DRS from Holdings Alpha} = \underbrace{\text{Dilution}}_{\approx 1\text{bp}} + \overbrace{\underbrace{\text{Price impact}}_{\approx 6\text{bp}} + \underbrace{\text{Liquidity tilt}}_{\approx 2\text{bp}}}^{\text{Alpha erosion}} \approx 9\text{bp}.$$

Our analysis also reveals heterogeneity — small funds with higher turnover experience stronger effects from price impact and liquidity tilts. If we focus on small-cap funds, the estimate for a typical such fund increases to around 13 basis points.

5 DRS in Return Gap

This section estimates the DRS from the return gap component as defined in equation (1). We estimate that for an average fund doubling its size, the DRS effect means a decline of 0.243% in annual returns. We find stronger DRS effects for funds with higher turnover and more illiquid holdings. For funds with the highest turnover and most illiquid holdings, we estimate that the DRS effect can be around 0.4%.

5.1 Average DRS effects in the return gap

To gauge the DRS in the return gap, we estimate a panel regression:

$$\text{ReturnGap}_{i,t} = -\,b \cdot \log_2(A_{i,t-1}) + \tau_t + \tau_i + \epsilon_{i,t} \tag{8}$$

where the dependent variable is the annualized fund return gap, τ_t and τ_i are month and fund fixed effects, and $b > 0$ is the DRS effect of doubling fund size (so the coefficient on $\log_2(A_{i,t-1})$ equals $-b$). We cluster standard errors by month and fund.

Table 4 reports the result. Column (1) delivers a full-sample estimate of -0.243 on $\log_2(A)$, which implies $b = 0.243$: when fund size doubles, the return gap component of fund return declines by 24.3 basis points. The statistical power is high as the standard error is only 0.032, and the t-statistic is 7.6. The high statistical power largely arises from the fact that the return gap has much lower variation than the holdings return, and thus the "signal-to-noise" ratio is high in this regression.

Dependent variable: annualized return gap (ReturnGap$_{i,t}$, in %)						
	(1)	(2)	(3)	(4)	(5)	(6)
$\log_2(A_{i,t-1})$	-0.243^{***}	-0.120^{***}	-0.163^{***}	-0.188^{***}	-0.066^{*}	-0.085^{**}
	(0.032)	(0.028)	(0.037)	(0.029)	(0.038)	(0.034)
$\log_2(A_{i,t-1}) \times$ Turnover		-0.257^{***}			-0.245^{***}	-0.247^{***}
		(0.060)			(0.059)	(0.058)
$\log_2(A_{i,t-1}) \times$ Stock Size (neg)			-0.152^{**}		-0.113^{*}	
			(0.065)		(0.063)	
$\log_2(A_{i,t-1}) \times$ BA Spread				-0.108^{**}		-0.080^{*}
				(0.050)		(0.048)
Month FE	Y	Y	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y	Y	Y
Obs	275,067	275,067	275,067	275,067	275,067	275,067
R^2	0.075	0.075	0.075	0.075	0.075	0.075
Within R^2	0.001	0.001	0.001	0.001	0.001	0.001

Table 4. The effect of fund size on return gap

The table presents panel regressions of annualized fund return gap in percent on lagged fund size transformed to base-2 logarithm. Due to this specification, the reported coefficients on $\log_2(A)$ equal $-b$, where $b > 0$ is the DRS effect when fund size doubles (equation (8)). Column (1) has no additional interactions. Columns (2)–(4) interact $\log_2(A)$ with the within-month percentile rank (Unif[0, 1]) of turnover, portfolio-weighted stock size (signed so that higher ranks correspond to smaller stocks), or portfolio-weighted bid–ask spread. Columns (5) and (6) include the turnover interaction together with the stock-size and bid–ask interactions, respectively. The table reports the $\log_2(A)$ coefficient and its interaction with the rank variable; the main effect of the rank variable is omitted. All regressions control for month and fund fixed effects. Standard errors are clustered by month and fund. The last row reports the R-squared not attributable to fixed effects. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

5.2 Mechanisms and heterogeneity

It is worth noting that the DRS arising from return gap appears much larger than that from the holdings alpha component. To better understand this DRS component, we delve into mechanisms, which also sheds light on DRS heterogeneity across funds.

What drives the return gap? For ease of reading, we reproduce the identity in equation (3), which shows that the return gap is given by the following decomposition:

$$\text{ReturnGap}_{i,t} = \underbrace{\sum_j \sum_{k=0}^{K-1} (w_{j,k} - w_{j,0}) \cdot \alpha_{j,k+1}}_{\text{weight deviation}} - \underbrace{\sum_j \sum_{k=0}^{K-1} w_{j,k} \text{TCost}_{j,k+1}}_{\text{trading costs}} + \text{Residual}$$

We argue that the DRS effect primarily arises from the first two terms, as the last "Residual" term is small. The residual term contains two small parts. One is net lending fees and financing costs: funds can lend out their shares to obtain lending fees; they also need to pay lending fees when shorting stocks. However, as discussed in Appendix B.1, these elements are small for active funds and have a negligible contribution to DRS. The other is approximation error arising from approximating the full-period fund alpha with the sum of subperiod alphas. Appendix B.2 investigates it empirically and also finds it to be a minuscule channel.

The first term, "weight deviation", summarizes all deviations of within-quarter portfolio weights from that observed from the quarterly 13F filings ($\{w_{j,0}\}$). We anticipate that the main part of it is driven by intraquarter trading where funds take advantage of fast-moving opportunities. Because the intraquarter portfolio weights are different from the quarter-end weights, the gap between fund return and holdings-implied returns widens. Consistent with this view, Appendix E shows that funds with higher turnover have more variable return gaps.[4]

The second term captures trading costs, which include trading commissions, bid-ask spreads, and temporary price impacts of trading.[5] Trading costs are a major component of the return gap: Edelen et al. (2013) estimate that for an average U.S. equity fund, annual trading costs are higher than the expense ratio.[6]

We hypothesize that both of the first two terms contribute to DRS effects. First, when funds become larger, they become less nimble at intraquarter trading, and thus the alpha they can earn from the first term declines. Consistent with this view, Appendix E finds that funds

[4]Note that portfolio weight deviations also include the new IPO allocations highlighted by Kacperczyk et al. (2008) as well as mismeasurements in 13F quarter-end holdings.

[5]In contrast, long-lasting price impact shows up as a component of holdings alpha, as discussed in Section 4.3.

[6]Their classification of transaction costs differs from ours. Specifically, Edelen et al. (2013) treat all price impact — around two-thirds of their total estimate — as a trading cost. In our decomposition, long-lasting price impact enters holdings alpha, while only temporary impact, commissions, and spreads enter the return gap. Even if we exclude the price-impact component of their estimate and retain only commissions and bid-ask spreads, the remaining drag is still sizeable.

reduce their turnover when becoming larger. Second, when funds become larger, their trades become a larger fraction of overall trading volume, which results in larger temporary price impacts which erodes alpha.

These mechanisms imply further heterogeneity of return gap-based DRS across funds, which we now test. The first effect implies that funds that are more active should experience stronger DRS effects. To test this, we turn the cross-sectional lagged turnover rank of funds into a Unif[0,1] variable and interact it with lagged fund size in the regression. The results are reported in Column (2) of Table 4. Consistent with our hypothesis, funds with higher turnover appear to suffer stronger DRS effects. Relative to funds with the lowest turnover, which experience a DRS effect of 0.120%, funds with the highest turnover experience a DRS effect of $0.120\% + 0.257\% = 0.377\%$, a point estimate that is approximately three times as high. The interaction coefficient is statistically significant at the 1% level.

The second effect implies that funds that trade less liquid stocks will experience larger DRS effects. To test this, we use two proxies of fund portfolio liquidity. The first is the one used in Section 4.3, the portfolio-weighted average of the size decile of the stocks it holds. We take the negative of this measure so that higher values mean *less* liquid portfolios. The second measure is the portfolio-weighted average of the bid-ask spread of the stocks held.[7] The regression results, reported in columns (3) and (4), are consistent with our hypothesis. Relative to funds with the most liquid holdings, those with the most illiquid holdings have DRS effects that are higher by 0.108% to 0.152%, and the interaction coefficients are statistically significant at the 5% level. These are not overlapping effects: the time-series averages of the monthly cross-sectional rank correlations of turnover with stock size and with bid-ask spreads are only -0.141 and 0.208, respectively. Consistent with the separation, columns (5) and (6) include both the turnover and liquidity interactions and find that both effects re-

[7] In the post-1993 period, we use the TAQ-based daily quoted spread summaries from WRDS intraday indicators and take monthly averages. For the pre-1993 period when TAQ is not available, we use the bid-ask spread estimates of Corwin and Schultz (2012); the data is downloaded from Professor Corwin's website.

main economically meaningful, even though the statistical significance drops somewhat for the liquidity interaction.

6 Implications

The previous two sections quantify DRS effects and conclude that, for an average fund, doubling fund size leads to a decline of annual alpha of around 0.33%. This section investigates the quantitative relevance of the BG (Berk and Green, 2004) framework for explaining stylized facts in active mutual funds.

6.1 Counterfactual performance without size adjustment

To investigate how much of the Berk–Green mechanism our estimated DRS can explain, we first define a counterfactual fund performance path if size had not adjusted over the previous h periods. We model fund alpha in a way similar to papers that estimate DRS statistically:

$$\alpha_{i,t} = \underbrace{a_i - b \cdot \log_2(A_{i,t-1})}_{E_{t-1}(\alpha_{i,t})} + \epsilon_{i,t},$$

where a_i is a time-invariant fund effect, b is the DRS slope, and $\epsilon_{i,t}$ is a mean zero i.i.d. noise shock.

$$\alpha_{i,t}^{(h)} = \underbrace{a_i - b \cdot \log_2(A_{i,t-h})}_{\text{skill}_{i,t}^{(h)}} + \underbrace{\epsilon_{i,t}}_{\text{luck}} \tag{9}$$

$$= \alpha_{i,t} + \underbrace{b \cdot \log_2\big(A_{i,t-1}/A_{i,t-h}\big)}_{\text{undoing DRS}}. \tag{10}$$

Mechanically, this adds back (subtracts) alpha for funds that grew (shrank) between $t-h$ and $t-1$: DRS converts past size changes into a return adjustment. Size changes include both net flows and the mechanical effect of returns on AUM; at the horizons we emphasize, flows

dominate. The subsections below apply $\alpha^{(h)}$ to performance persistence (with h matched to the AR horizon) and to the skill share of return variance (where longer h is also informative).

6.2 Performance persistence without DRS

A central implication of Berk and Green (2004) is that, because flows chase performance, DRS effects erase performance persistence. We quantify the magnitude of the DRS effect on the AR(1) coefficient of annual fund alphas, a quantitative question also highlighted by Reuter and Zitzewitz (2021).

We first estimate the AR(1) coefficient of annual fund alphas using Fama–MacBeth regressions in our non-sector sample. The results are reported in Panel A of Table 5. Depending on the model specification, we find a coefficient of 0.055 (Morningstar benchmark-adjusted return, insignificant) to 0.117 (FF5 alphas, $t = 2.91$).

How much of this low persistence is due to size adjustment? Under the model in Section 6.1, $\alpha_{i,y} = a_i - b \cdot \log_2(A_{i,y-1}) + \epsilon_{i,y}$, performance-chasing size growth lowers next year's alpha by b times that growth, which biases the observed AR(1) downward by

$$b \times \underbrace{\frac{\mathrm{Cov}\big(\log_2(A_{i,y}/A_{i,y-1}), \alpha_{i,y}\big)}{\mathrm{Var}(\alpha_{i,y})}}_{\beta^{\text{size}} \equiv \text{size-on-alpha slope}}.$$

Panel B of Table 5 estimates this size-on-alpha slope β^{size} and implies that a one-percentage-point higher annual alpha is associated with about 1.5% larger fund size. Panel C reports the product with our DRS slope of $b = 0.0033$. We find that DRS effects can reduce the AR(1) coefficient of annual fund alphas by about 0.007.

DRS at other levels Our paper focuses on estimating fund-level DRS. When applying this to understanding the impact on performance persistence, however, one may be concerned about DRS operating at other levels. Pástor and Stambaugh (2012) and Pástor et al. (2015)

Alpha model:	(1) benchmark-adj	(2) CAPM	(3) FF3	(4) Carhart	(5) FF5	(6) FF6
Panel A: Annual AR(1) of fund alpha						
	Dependent variable: $\alpha_{i,y}$					
$\alpha_{i,y-1}$	0.055	0.130	0.084*	0.075*	0.117***	0.088**
	(0.041)	(0.083)	(0.043)	(0.041)	(0.040)	(0.034)
Obs	9,083	9,083	9,083	9,083	9,083	9,083
Avg. R^2	0.045	0.148	0.046	0.045	0.046	0.031
Panel B: Size growth on annual alpha						
	Dependent variable: $\log_2(A_{i,y}/A_{i,y-1})$					
$\alpha_{i,y}$	2.064***	2.080***	2.217***	2.254***	2.143***	2.257***
	(0.150)	(0.137)	(0.138)	(0.133)	(0.145)	(0.140)
Obs	9,083	9,083	9,083	9,083	9,083	9,083
Avg. R^2	0.137	0.205	0.145	0.145	0.129	0.132
Panel C: Implied DRS effect on AR(1) coef						
$b \times \hat{\beta}^{\text{size}}$	0.007	0.007	0.007	0.007	0.007	0.007

Table 5. Quantifying the DRS effect on annual alpha persistence

Estimates in Panels A and B are Fama–MacBeth averages of annual cross-sectional regressions. Columns vary the alpha model. Panel A reports the AR(1) coefficient of annual fund alpha. Panel B reports the slope of year-over-year $\log_2$ AUM growth on annual alpha. Panel C reports the implied one-year downward bias on the AR(1) coefficient, $b \times \hat{\beta}^{\text{size}}$, using our estimated DRS of $b = 0.0033$ (0.33% of annual alpha per doubling of fund size) treated as known. Standard errors in Panels A and B are Fama–MacBeth. Levels of significance: *p<0.1; **p<0.05; ***p<0.01.

emphasize industry-level decreasing returns: active funds earn higher alphas when the *entire* active sector is smaller. For our persistence exercise, such industry-wide scale effects are largely a common factor. Cross-sectional AR(1) coefficients compare relative performance across funds, so a shock that compresses alphas for all active managers alike does not by itself explain why past winners fail to remain winners.

A sharper concern is peer-group DRS: if funds that pursue overlapping ideas experience correlated flows, size growth at peers could erode a winner's future alpha beyond own-size DRS. Appendix F.1 investigates this by using a holdings-based style distance following Hoberg et al. (2018) and a holdings-overlap distance in the spirit of Wahal and Wang (2011) to find the peer group for each fund. We then control for peer-group AUM in DRS regressions. Own-size DRS effects are essentially unchanged, and we find little evidence for peer fund

DRS.

6.3 Skill share of return variance without DRS

Another debate in the literature is the share of realized fund performance that can be attributed to "skill" or "luck". The traditional view is that performance is mostly luck, and thus investors chasing past winners is futile (e.g. Fama and French, 2010). BG also predict that fund performance is unpredictable, but argue that this arises from DRS effects. If there were no DRS effects and fund size changes do not erase fund skill, then a higher fraction of fund performance would have been predictable and attributed to skill.

Under the model in Section 6.1, write realized performance as

$$\alpha_{i,t} = \underbrace{a_i - b \cdot \log_2(A_{i,t-1})}_{\text{skill}_{i,t} \equiv E_{t-1}(\alpha_{i,t})} + \underbrace{\epsilon_{i,t}}_{\text{luck}}. \tag{11}$$

Following Fama and French (2010), "skill" is defined as the predictable / expected component of alpha. The skill-to-performance ratio, $\text{Var}(\text{skill}_{i,t})/\text{Var}(\alpha_{i,t})$, is related to the AR(1) persistence of fund alphas. Note:

$$\phi = \frac{\text{Cov}(\alpha_{i,t}, \alpha_{i,t-1})}{\text{Var}(\alpha_{i,t-1})}.$$

Rewrite next-period alpha by inserting last period's skill and the intervening size change,

$$\begin{aligned}\alpha_{i,t} &= a_i - b \cdot \log_2(A_{i,t-2}) - b \cdot \big(\log_2(A_{i,t-1}) - \log_2(A_{i,t-2})\big) + \epsilon_{i,t} \\ &= \text{skill}_{i,t-1} - b \cdot \Delta \log_2(A_{i,t-1}) + \epsilon_{i,t}.\end{aligned}$$

Taking the covariance with $\alpha_{i,t-1}$ and using $\alpha_{i,t-1} = \text{skill}_{i,t-1} + \epsilon_{i,t-1}$, with $\epsilon_{i,t}$ orthogonal to

date-$t-1$ objects and luck orthogonal to predetermined skill,

$$\mathrm{Cov}(\alpha_{i,t}, \alpha_{i,t-1}) = \mathrm{Var}(\mathrm{skill}_{i,t-1}) - b \cdot \mathrm{Cov}\big(\Delta \log_2(A_{i,t-1}), \alpha_{i,t-1}\big).$$

Dividing by $\mathrm{Var}(\alpha_{i,t-1})$ yields

$$\phi = \underbrace{\frac{\mathrm{Var}(\mathrm{skill}_{i,t-1})}{\mathrm{Var}(\alpha_{i,t-1})}}_{\text{skill-to-performance ratio}} \underbrace{- b \cdot \frac{\mathrm{Cov}\big(\Delta \log_2(A_{i,t-1}), \alpha_{i,t-1}\big)}{\mathrm{Var}(\alpha_{i,t-1})}}_{\text{DRS adjustment from return chasing}}.$$

The fraction in the second term is the size-on-alpha slope β^{size} from Section 6.2, so the skill fraction equals $\phi + b\beta^{\text{size}}$. Note that DRS effects form the key wedge between the AR(1) coefficient ϕ — which is a modest number in practice — and the skill-to-performance ratio. If DRS effects are larger, then the skill-to-performance ratio can be high even if ϕ is low. However, because b is small ($b\beta^{\text{size}} \approx 0.007$), ϕ is already a tight lower bound on the skill share. Based on the estimates in Table 5, the fraction of skill in annual performance ranges from around $0.055 + 0.007 = 0.062$ when using Morningstar benchmark-adjusted return to $0.130 + 0.007 = 0.137$ when using CAPM alpha as performance measures.

A natural BG-based response is that observed ϕ is low precisely because size adjusts: winners grow, DRS erodes their edge, and persistence is compressed. If size were fixed and did not respond to performance, persistence — and thus the skill share — would be higher. To quantify that counterfactual, freeze size h years earlier using $\alpha^{(h)}$ in (10) (with $b = 0.0033$). Let $\phi^{(h)}$ denote the Fama–MacBeth AR(1) coefficient of $\alpha^{(h)}$ — our measure of *counterfactual persistence* when size is held fixed for h years. Because the DRS channel from intervening size change has been shut down inside $\alpha^{(h)}$, no further $+b\beta^{\text{size}}$ adjustment is needed: $\phi^{(h)}$ itself is the skill-share proxy. For notational uniformity we also write $\mathrm{skill}^{(0)}_{i,t} \equiv \mathrm{skill}_{i,t}$, $\alpha^{(0)}_{i,t} \equiv \alpha_{i,t}$, and $\phi^{(0)} \equiv \phi$ (observed persistence).

Table 6 reports the results. Panel B at $h = 0$ puts the skill fraction at about 0.06–0.14

Alpha model:	(1) benchmark-adj	(2) CAPM	(3) FF3	(4) Carhart	(5) FF5	(6) FF6
	Panel A: Counterfactual persistence $\phi^{(h)}$					
$h = 0$	0.055	0.130	0.084	0.075	0.117	0.088
$h = 1$	0.068	0.151	0.083	0.080	0.101	0.081
$h = 2$	0.093	0.142	0.100	0.097	0.115	0.100
$h = 3$	0.058	0.113	0.106	0.093	0.103	0.089
$h = 5$	0.036	0.131	0.085	0.080	0.090	0.088
$h = 7$	0.020	0.062	0.066	0.059	0.082	0.077
$h = 10$	0.058	0.025	−0.018	−0.021	0.036	0.016
Max	0.093	0.151	0.106	0.097	0.117	0.100
	Panel B: Fraction of alpha variance explained by skill					
$h = 0$	0.062	0.137	0.091	0.083	0.124	0.095
$h = 1$	0.068	0.151	0.083	0.080	0.101	0.081
$h = 2$	0.093	0.142	0.100	0.097	0.115	0.100
$h = 3$	0.058	0.113	0.106	0.093	0.103	0.089
$h = 5$	0.036	0.131	0.085	0.080	0.090	0.088
$h = 7$	0.020	0.062	0.066	0.059	0.082	0.077
$h = 10$	0.058	0.025	−0.018	−0.021	0.036	0.016
Max	0.093	0.151	0.106	0.097	0.124	0.100

Table 6. Skill share of annual alpha variance

Panel A reports counterfactual persistence $\phi^{(h)}$, the Fama–MacBeth AR(1) of annual $\alpha^{(h)}$. For $h \geq 1$, $\alpha^{(h)}$ is the size-frozen counterfactual in (10): theoretically $a_i - b \log_2(A_{i,t-h}) + \epsilon_{i,t}$, and computed from the data as $\alpha_{i,t} + b \log_2(A_{i,t-1}/A_{i,t-h})$ with $b = 0.0033$. The row $h = 0$ is observed persistence ϕ ($\alpha^{(0)} \equiv \alpha$). Panel B reports the implied fraction of cross-sectional alpha variance explained by skill: $\phi + b\beta^{\text{size}}$ at $h = 0$, and counterfactual persistence $\phi^{(h)}$ for $h \geq 1$. Columns vary the alpha model. The sample is non-sector U.S. active equity funds. Max is the largest entry across the listed horizons.

across models ($\phi + b\beta^{\text{size}}$), so skill accounts for on the order of 10% of cross-sectional alpha variance. If the cross-sectional standard deviation of annualized alpha is around 6%, the implied skill standard deviation is about $\sqrt{0.12} \times 6\% \approx 2\%$, which is in the ballpark of Fama and French (2010). Counterfactual persistence does not overturn this conclusion: the largest $\phi^{(h)}$ across models and $h \in \{0, 1, 2, 3, 5, 7, 10\}$ is about 0.15 (CAPM); estimates are not monotone in h because long-horizon size paths mix performance chasing with other non-performance size variation. Even after shutting down many years of size adjustment, counterfactual persistence — and thus the skill share — remains a small fraction of alpha variance. A raw AR(1) that ignores DRS is, for our estimate of DRS b, already close to the right skill-share bound.

7 Conclusion

Decreasing returns to scale is a central parameter for differentiating the traditional and Berk–Green (BG) views of active funds. Existing studies estimate the reduced-form relationship between size and performance and the resulting magnitudes span a wide range. This paper offers a complementary, bottom-up approach grounded in holdings and performance data. For a typical active U.S. equity mutual fund, we find that DRS is modest — on the order of 0.3 to 0.4 percentage points of annual alpha when fund size doubles.

Our method also sheds light on the relative importance of different components of DRS. We use an accounting identity to decompose fund alpha into holdings alpha — the return implied by beginning-of-period positions — and the Kacperczyk et al. (2008) return gap. Holdings data discipline the first component: we observe how 13F quarterly portfolio holdings change as funds scale up and translate those changes into the impact on fund returns. We estimate that DRS in holdings alpha is small — about 9 bp of annual alpha per size doubling for a typical fund (Section 4.4). In contrast, the second term — the return gap — which includes intraquarter trading, transaction costs, and related frictions — is the main source of DRS, contributing about 24 bp per doubling. This is consistent with the idea that when funds become larger, they become less nimble at intraquarter trading and also incur larger transaction costs. Adding the two components yields a total effect of roughly 0.33% for the average fund.

DRS is heterogeneous. Effects are larger for high-turnover funds and those holding less liquid stocks — in both holdings alpha and the return gap — and can reach about 0.5% of annual alpha per size doubling.

We end by quantifying what these magnitudes imply for BG's quantitative account of mutual fund facts. The mechanism is the right one in principle: performance-chasing size adjustment can compress persistence and mask skill. With our estimated DRS magnitudes, however,

the effects are limited for U.S. active equity funds. For instance, for explaining the lack of performance persistence, we find that undoing DRS effects raises the AR(1) coefficient of annual fund alpha by only around 0.007. Similarly, assuming that fund sizes do not adjust for up to ten years has limited impact on the share of "skill" in cross-sectional alpha variation. Our findings suggest that while DRS effects do exist, their quantitative explanatory power for mutual funds in the stock market — which is relatively liquid — is limited. We expect BG to have larger quantitative explanatory power in less liquid settings — including venture capital, private equity, and hedge funds that trade more frequently or hold assets that are less liquid than equities.

References

Adams, J. C., D. K. Hayunga, and S. Mansi. 2018a. Diseconomies of Scale in the Actively-Managed Mutual Fund Industry: What Do the Outliers in the Data Tell Us? *Critical Finance Review, Forthcoming* .

Adams, J. C., D. K. Hayunga, and S. Mansi. 2018b. Scale and performance in active management are not negatively related. *Critical Finance Review, forthcoming* .

Anton, M., R. B. Cohen, and C. Polk. 2021. Best ideas. *Available at SSRN 1364827* .

Barras, L., P. Gagliardini, and O. Scaillet. 2022. Skill, scale, and value creation in the mutual fund industry. *The Journal of Finance* 77:601–638.

Berk, J. B., and R. C. Green. 2004. Mutual fund flows and performance in rational markets. *Journal of political economy* 112:1269–1295.

Berk, J. B., and J. H. Van Binsbergen. 2015. Measuring skill in the mutual fund industry. *Journal of financial economics* 118:1–20.

Bershova, N., and D. Rakhlin. 2013. The Non-Linear Market Impact of Large Trades: Evidence from Buy-Side Order Flow. *Quantitative Finance* 13:1759–1778.

Busse, J. A., T. Chordia, L. Jiang, and Y. Tang. 2021. Transaction costs, portfolio characteristics, and mutual fund performance. *Management Science* 67:1227–1248.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of finance* 52:57–82.

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.

Chevalier, J., and G. Ellison. 1997. Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy* 105:1167–1200.

Corwin, S. A., and P. Schultz. 2012. A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance* 67:719–760.

Coval, J., and E. Stafford. 2007. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics* 86:479–512.

Davis, C., X. Han, S. Sokolinski, and A. Tamoni. 2026. Erasing Alpha .

Edelen, R., R. Evans, and G. Kadlec. 2013. Shedding light on "invisible" costs: Trading costs and mutual fund performance. *Financial Analysts Journal* 69:33–44.

Evans, R. B., M. Rohleder, H. Tentesch, and M. Wilkens. 2023. Diseconomies of Scale in Quantitative and Fundamental Investment Styles. *Journal of Financial and Quantitative Analysis* 58:2417–2445.

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

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* 17:483–525.

Frazzini, A., R. Israel, and T. J. Moskowitz. 2018. Trading costs. *Available at SSRN 3229719* .

Harvey, C. R., and Y. Liu. 2021. Decreasing returns to scale, fund flows, and performance. *Fund Flows, and Performance (June 21, 2021)* .

Hoberg, G., N. Kumar, and N. Prabhala. 2018. Mutual fund competition, managerial skill, and alpha persistence. *The Review of Financial Studies* 31:1896–1929.

Honkanen, P. 2025. Securities lending and trading by active and passive funds. *Journal of Financial and Quantitative Analysis* 60:1272–1309.

Huang, S., X. Lu, Y. Song, and H. Xiang. 2023. Remeasuring Scale in Active Management. *Available at SSRN* .

Jensen, M. C. 1968. The performance of mutual funds in the period 1945-1964. *The Journal of finance* 23:389–416.

Kacperczyk, M., C. Sialm, and L. Zheng. 2008. Unobserved actions of mutual funds. *The Review of Financial Studies* 21:2379–2416.

Lou, D. 2012. A flow-based explanation for return predictability. *The Review of Financial Studies* 25:3457–3489.

Malkiel, B. G. 1995. Returns from investing in equity mutual funds 1971 to 1991. *The Journal of finance* 50:549–572.

Pástor, L., and R. F. Stambaugh. 2012. On the size of the active management industry. *Journal of Political Economy* 120:740–781.

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

Pástor, L., R. F. Stambaugh, and L. A. Taylor. 2017. Do funds make more when they trade more? *The Journal of Finance* 72:1483–1528.

Pástor, L., R. F. Stambaugh, and L. A. Taylor. 2020. Fund tradeoffs. *Journal of Financial Economics* 138:614–634.

Pástor, L., R. F. Stambaugh, L. A. Taylor, M. Zhu, et al. 2022. Diseconomies of Scale in Active Management: Robust Evidence. *Critical Finance Review* 11:593–611.

Perold, A. F., and R. S. Salomon. 1991. The Right Amount of Assets Under Management. *Financial Analysts Journal* 47:31–39.

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 & Finance* 88:357–365.

Pollet, J. M., and M. Wilson. 2008. How does size affect mutual fund behavior? *The Journal of Finance* 63:2941–2969.

Reuter, J., and E. Zitzewitz. 2021. How much does size erode mutual fund performance? A regression discontinuity approach. *Review of Finance* 25:1395–1432.

Rizova, S. 2011. Securities lending by mutual funds. *Available at SSRN 2056651* .

Shumway, T., M. Szefler, and K. Yuan. 2009. The information content of revealed beliefs in portfolio holdings. *Available at SSRN 1364733* .

Sirri, E. R., and P. Tufano. 1998. Costly Search and Mutual Fund Flows. *Journal of Finance* 53:1589–1622.

Van Binsbergen, J. H., J. H. J. Kim, and S. Kim. 2025. Capital allocation and the market for mutual funds: Inspecting the mechanism. *Journal of Financial and Quantitative Analysis* .

Wahal, S., and A. Y. Wang. 2011. Competition among mutual funds. *Journal of Financial Economics* 99:40–59.

Wermers, R. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance* 55:1655–1695.

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

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

A DRS estimates in the literature

This section surveys DRS estimates in the literature and converts them to a common unit of measurement: the decline in annual alpha of a typical active U.S. equity fund if its size exogenously doubles. Table A.1 summarizes the converted estimates. Figure A.1 plots the high-end and low-end estimates; we omit the insignificant estimates because their large standard errors would dominate the scale of the figure.

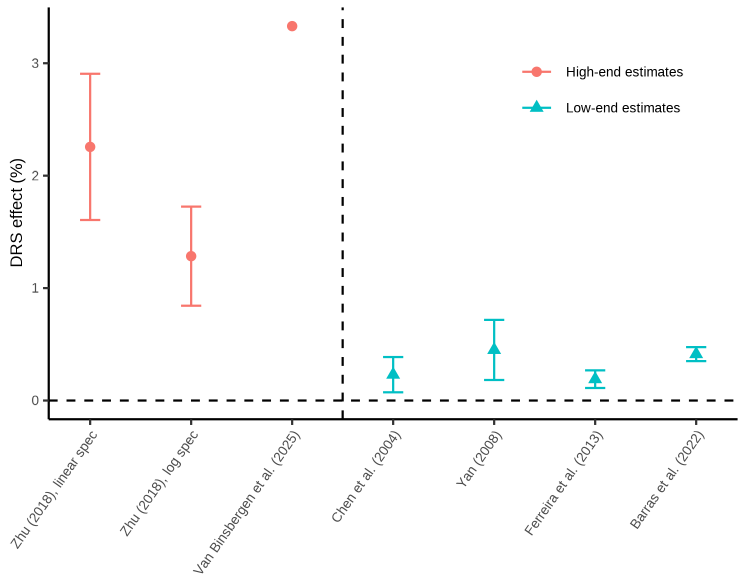


Figure A.1. DRS effect estimates in the literature

This figure plots the high-end (red circles) and low-end (blue triangles) DRS estimates from Table A.1. The vertical axis is the decline in annual alpha in percentage points if the size of a typical active U.S. equity fund doubles. Vertical bars are 95% confidence intervals; Van Binsbergen et al. (2025), plotted in the third data point, does not report a standard error for the average effect.

Below we describe how we extract the estimates in these studies to convert to our common unit for DRS effects.

High-end estimates

- **Zhu (2018)** develops an econometric improvement on the recursive demean estimator of Pástor et al. (2015) for estimating DRS. The paper specifies two specifications, one where fund alpha is linear in fund size and one where it is linear in log size. The paper also reports large heterogeneity in DRS estimates across funds with different sizes.
  To obtain estimates for a typical fund, we take the estimates for a median-sized fund. In their Table 4's group 6 whose fund size of $270 million is close to the median fund size of $261 million in their sample (see their Table 1), Zhu (2018) reports that doubling

Estimate magnitude	Paper	DRS effect (%)	Standard error (%)
(1)	(2)	(3)	(4)
High-end	Zhu (2018), linear specification	2.26	0.33
	Zhu (2018), log specification	1.28	0.22
	Van Binsbergen et al. (2025)	3.33	N/A
Low-end	Chen et al. (2004)	0.23	0.08
	Yan (2008)	0.45	0.14
	Ferreira et al. (2013)	0.19	0.04
	Barras et al. (2022)	0.41	0.03
Insignificant	Phillips et al. (2018)	5.5	4.4
	Reuter and Zitzewitz (2021)	1.66	3.35

Table A.1. Decreasing returns to scale (DRS) estimates in the literature

This table summarizes existing estimates of fund-level DRS for active U.S. equity mutual funds, converted to a common metric: the decline in annual alpha (in percentage points) if the size of a typical active fund doubles. For instance, the estimate in the first row means if a typical fund's size doubles, its annual alpha declines by 2.26% (e.g. from 3% to 0.74%). Column (1) groups estimates by magnitude. Column (2) lists the source paper (and specification, where relevant). Columns (3) and (4) report the implied DRS effect and its standard error when applicable. High-end and low-end estimates are statistically distinguishable from zero; the insignificant estimates come from designs that exploit exogenous shocks to fund size.

fund sizes leads annual alpha to decline by $12 \times 0.188\% = 2.256\%$ for the linear model and by $12 \times 0.107\% = 1.284\%$ for the log model. The standard errors can be computed using the reported t-statistics.

- **Van Binsbergen et al. (2025)** estimate fund-specific DRS parameters from fund-by-fund time-series regressions of monthly alpha on lagged log real AUM with a small-sample bias correction. In their section 4.A, they report that, for the average fund, a 1% increase in fund size is on average associated with a decline in CAPM alpha of about 0.4 basis points per month (see their discussion of Figure 1). Under their log-size specification, doubling fund size raises log size by $\ln 2$, so the implied annual alpha decline is $12 \times (0.4/100) \times \ln 2 \approx 3.33\%$. Because the paper focuses on cross-fund heterogeneity in DRS, it does not report a standard error for this average effect.

Low-end estimates

- **Chen et al. (2004)** is the first paper that estimates DRS effects. They run Fama–MacBeth regressions of monthly benchmark-adjusted fund returns on lagged log fund size (natural log of TNA), controlling for family size, turnover, age, fees, loads, flows, and lagged returns, and they exclude the smallest fund-size quintile from the baseline sample.

  In their Table 3, the coefficient on lagged log TNA for market-adjusted gross returns

is -0.028 (Newey–West $t = 3.02$), or about -2.8 basis points per month per unit of log size. Under their log-size specification, doubling fund size raises log size by $\ln 2$, so the implied annual alpha decline is $12 \times 0.028 \times \ln 2 \approx 0.23\%$, with a standard error of $12 \times (0.028/3.02) \times \ln 2 \approx 0.08\%$.

- **Yan (2008)** follow Chen et al. (2004) and run Fama–MacBeth regressions of monthly factor alphas on lagged log TNA with controls. In their Table 8, the log-TNA coefficient for four-factor gross alphas is -0.054 ($t = -3.30$), so doubling size implies an annual alpha decline of $12 \times 0.054 \times \ln 2 \approx 0.45\%$ (SE $\approx 0.14\%$).
- **Ferreira et al. (2013)** run panel regressions of quarterly Carhart four-factor alphas on lagged log TNA for domestic equity funds in 2000–2007. For U.S. funds (their Table V, column 1), the log-TNA coefficient is -0.0675 ($t = -4.99$), so doubling size implies an annual alpha decline of $4 \times 0.0675 \times \ln 2 \approx 0.19\%$ (SE $\approx 0.04\%$).
- **Barras et al. (2022)** estimate fund-specific skill and scale coefficients from the linear specification $\alpha_{i,t} = a_i - b_i q_{i,t-1}$, where $q_{i,t-1}$ is lagged real AUM, using a bias-adjusted estimator that corrects for the endogeneity of size. In Panel B of their Table II, in the full sample estimate ("All funds"), they find that increasing fund size by one standard deviation ($655 million from their footnote 19) leads to a decline of annual gross alpha by 1.3% with a standard error of 0.1%. Thus, for the typical fund which has size $208 million in their sample (the Panel C of their Table I), doubling fund size leads to a decline of annual alpha by $1.3\% \times (208/655) \approx 0.41\%$ with a standard error of $0.1\% \times (208/655) \approx 0.03\%$.

  The paper finds DRS to be heterogeneous across funds. In Panel B of their Table II, they estimate DRS for funds with different capitalization, turnover, and broker-versus-direct sold characteristics. When applying the same calculation, their results imply a range of DRS effects from 0.30% for low-turnover funds to 0.57% for broker-sold funds.

Insignificant estimates The next two papers estimate DRS from exogenous shocks to fund sizes. Those shocks are much smaller than a doubling of assets, so converting their estimates into our common metric requires extrapolating beyond the variation used for identification.

- **Phillips et al. (2018)** address endogeneity with an instrumental variable strategy based on fund-flow reactions to mechanical changes in reported past holding-period returns (HPRs): when old returns drop out of the lookback window, advertised HPRs change and flows respond even though the signal is stale. In annual 2SLS regressions (their Table 2, Panel B), the coefficient on predicted log fund size from the first stage

for annual decimal market-adjusted net returns is -0.08 ($t = 1.27$). Thus, this estimate implies that doubling fund size is associated with an annual alpha change of $0.08 \times \ln 2 \approx 5.5\%$ with a standard error of $(0.08/1.27) \times \ln 2 \approx 4.4\%$. This is statistically indistinguishable from zero.

- **Reuter and Zitzewitz (2021)** address endogeneity with a regression discontinuity design that exploits discrete jumps in Morningstar ratings around the four- to five-star threshold: funds whose past risk-adjusted returns fall just above the cutoff receive a higher star rating and significantly larger inflows. They estimate the causal effect of these star-induced flows on future returns via 2SLS (their Table VIII). For actively managed equity funds, the IV estimate for cumulative log net returns over the subsequent 12 months —expressed in the same units as the OLS coefficient on log portfolio size—is -0.0240 with standard error 0.0483. Thus, extrapolating to a doubling of fund size, this estimate implies an annual alpha decline of $0.0240 \times \ln 2 \approx 1.66\%$ with a standard error of $0.0483 \times \ln 2 \approx 3.35\%$, which is statistically indistinguishable from zero.

B Minor influences on DRS

This section discusses minor influences on DRS effects.

B.1 Net lending fees and financing costs

Net lending fees are small for active mutual funds. Our main sample already excludes funds that report short positions, which rules out lending fee expenses. As for lending income, unlike passive funds who have stable holdings and can lend out shares over long periods of time, active funds are less engaged in securities lending. Consistent with this, Honkanen (2025) find that the average annual net lending fee is only around 0.01% for active domestic equity funds. Prior literature also found that lending fees have limited variation with fund size. Table IV in Rizova (2011) regresses abnormal value added from securities lending — net lending income over AUM in excess of the asset-class–year median — on lagged log fund assets in a pooled cross section with year fixed effects. The coefficient on log(Assets) is about -0.42, which implies that when fund size doubles, abnormal lending income declines by only about $0.42 \times \log(2) \approx 0.29$ bp, or $just 0.0029\%$, per year — economically tiny relative to the return-gap DRS magnitudes in the main text.

B.2 Approximation error

The accounting identities in equations (2)–(3) are approximate because simple returns do not aggregate additively over time: $(1+r_1)(1+r_2) \neq 1+r_1+r_2$. We clarify the two compounding pieces that make up this approximation error and then estimate their contribution to DRS.

Mechanism. There are two pieces of approximation error. First, recall that each subperiod fund alpha is exact,

$$\alpha_{k+1}^{\text{fund}} = \sum_j w_{j,k}\big(\alpha_{j,k+1} - \text{TCost}_{j,k+1}\big),$$

but equation (2) replaces the true full-period fund alpha with the *sum* of these subperiod alphas. When full-period alpha is defined as the gap between compounded fund and benchmark simple returns, $\alpha_t^{\text{fund}} = R^{\text{fund}} - R^{\text{bench}}$ with $R^x = \prod_k (1 + R_{k+1}^x) - 1$, the associated compounding error is

$$\begin{aligned} \mathcal{E}_{\text{comp}} &\equiv \alpha_t^{\text{fund}} - \sum_{k=0}^{K-1} \alpha_{k+1}^{\text{fund}} = \mathcal{E}_{\text{comp}}^{\text{fund}} - \mathcal{E}_{\text{comp}}^{\text{bench}}, \\ \mathcal{E}_{\text{comp}}^{x} &= \left[\prod_{k=0}^{K-1}\big(1 + R_{k+1}^x\big) - 1\right] - \sum_{k=0}^{K-1} R_{k+1}^x, \qquad x \in \{\text{fund}, \text{bench}\}. \end{aligned} \tag{A.1}$$

The benchmark piece enters the *level* of $\mathcal{E}_{\text{comp}}$, but it does not depend on the fund's own AUM, and thus is irrelevant for DRS. Equivalently, with style-by-period fixed effects in regressions, $\mathcal{E}_{\text{comp}}^{\text{bench}}$ is absorbed. A second-order expansion yields $\mathcal{E}_{\text{comp}}^{x} \approx -\frac{1}{2}\sum_k (R_{k+1}^x)^2$, so the error scales with return variance.

A second piece arises when holdings alpha uses the full-period security alpha α_j while the $w_{j,0}$ peel in (3) uses $\sum_k \alpha_{j,k+1}$:

$$\mathcal{E}_{\text{sec}} = \sum_j w_{j,0}\left(\alpha_j - \sum_{k=0}^{K-1} \alpha_{j,k+1}\right). \tag{A.2}$$

If each α_j is likewise excess of a compounded benchmark, then $\mathcal{E}_{\text{sec}} = \sum_j w_{j,0}\,\mathcal{E}_{\text{comp}}^{j} - \mathcal{E}_{\text{comp}}^{\text{bench}}$, and again the pure benchmark term does not vary with fund size. Peeling the identity then yields that full-period fund alpha equals holdings alpha plus weight deviation minus trading costs plus $(\mathcal{E}_{\text{comp}} - \mathcal{E}_{\text{sec}})$, so the compounding residual that enters the return gap is

$$\text{ApproximationError} = \mathcal{E}_{\text{comp}} - \mathcal{E}_{\text{sec}}. \tag{A.3}$$

(The benchmark pieces cancel in this difference, leaving $\mathcal{E}^{\text{fund}}_{\text{comp}} - \sum_j w_{j,0}\,\mathcal{E}^{j}_{\text{comp}}$.)

Estimation. We construct empirical analogues of $\mathcal{E}_{\text{comp}}$ and $\mathcal{E}_{\text{sec}}$ in equation (A.3) at the fund-quarter frequency and regress each — as well as ApproximationError $= \mathcal{E}_{\text{comp}} - \mathcal{E}_{\text{sec}}$ — on lagged $\log_2(A)$, with style-by-quarter and fund fixed effects and two-way clustering by quarter and fund. (Relative to the main-text return-gap design, we replace calendar-time fixed effects with style-by-quarter fixed effects so that style-level benchmark compounding is absorbed.) Because the benchmark component drops out of the size slope as discussed earlier, for $\mathcal{E}_{\text{comp}}$ it suffices to use daily CRSP fund returns and compute the fund compound-minus-sum gap $\left[\prod_\tau (1+r_\tau) - 1\right] - \sum_\tau r_\tau$. For $\mathcal{E}_{\text{sec}}$, we compute the analogous compound-minus-sum gap for daily security alphas, aggregate to the quarter, and value-weight by beginning-of-quarter 13F holdings. Annualizing, the fund compounding error averages only about -0.18% per year in levels. Its within-fund size slope is about -0.1 basis points when fund size doubles (standard error ≈ 3 bp), and is statistically indistinguishable from zero. For ApproximationError, the point estimate remains on the order of a few basis points and is typically insignificant (e.g., about $+0.6$ bp for FF3 alphas). Overall, we conclude that the approximation error is a small channel for DRS.

C Analytical derivations

C.1 Fund alpha decomposition

This section provides the derivation of the fund alpha decomposition in Section 2.1. To simplify notation, we do not discuss non-portfolio-based income such as lending fees, etc. Those fees all enter into the residual term.

Consider a fund evaluated over a time period $[0, 1]$ (e.g., a quarter), which is partitioned into K subperiods. Let $k = 0, \ldots, K-1$ index the start of each subperiod. Let $w_{j,k}$ denote the portfolio weight of security j at time k, where $j = 1, \ldots, N$.

We decompose the simple return of each security into its benchmark component and its alpha: $R_{j,k+1} = \beta_j R^{\text{bench}}_{k+1} + \alpha_{j,k+1}$. We write a one-factor model for simplicity, but all linear multi-factor models can be equivalently expressed as a one-factor model where the factor is the mean-variance efficient combination of all factors. The factor loading of the fund in subperiod $k+1$ is given by $\beta^{\text{fund}}_{k+1} = \sum_{j=1}^{N} w_{j,k}\beta_j$.

The fund's gross simple alpha in subperiod $k+1$ is the weight-averaged alpha of its holdings,

net of any trading costs ($\text{TCost}_{j,k+1}$) incurred to adjust the portfolio:

$$\alpha_{k+1}^{\text{fund}} = \sum_{j=1}^{N} w_{j,k} \left(\alpha_{j,k+1} - \text{TCost}_{j,k+1}\right) \tag{A.4}$$

To aggregate these intra-period alphas into the total alpha over the full period $[0, 1]$, we must account for geometric compounding. The true cumulative fund return is the product of subperiod returns: $1 + R^{\text{fund}} = \prod_{k=0}^{K-1}(1 + R_{k+1}^{\text{fund}})$. The arithmetic sum of subperiod alphas is an approximation of the true compounded alpha. We define the geometric compounding error, which scales with the variance of the returns, as $\mathcal{E}_{\text{comp}}$:

$$\alpha_t^{\text{fund}} = \sum_{k=0}^{K-1} \alpha_{k+1}^{\text{fund}} + \mathcal{E}_{\text{comp}} \tag{A.5}$$

Substituting Equation (A.4) into the sum, we get:

$$\alpha_t^{\text{fund}} = \sum_{k=0}^{K-1} \sum_{j=1}^{N} w_{j,k} \alpha_{j,k+1} - \sum_{k=0}^{K-1} \sum_{j=1}^{N} w_{j,k} \text{TCost}_{j,k+1} + \mathcal{E}_{\text{comp}} \tag{A.6}$$

Next, we rewrite the dynamic weight $w_{j,k}$ as the initial weight plus the intra-period weight deviation: $w_{j,k} = w_{j,0} + (w_{j,k} - w_{j,0})$. Distributing this through the summation yields:

$$\alpha_t^{\text{fund}} = \sum_{j=1}^{N} w_{j,0} \left(\sum_{k=0}^{K-1} \alpha_{j,k+1} \right) + \sum_{j=1}^{N} \sum_{k=0}^{K-1} (w_{j,k} - w_{j,0}) \alpha_{j,k+1} - \sum_{j=1}^{N} \sum_{k=0}^{K-1} w_{j,k} \text{TCost}_{j,k+1} + \mathcal{E}_{\text{comp}} \tag{A.7}$$

Finally, recognizing that the arithmetic sum of security j's subperiod alphas is an approximation of its total period alpha, $\alpha_j \approx \sum_{k=0}^{K-1} \alpha_{j,k+1}$, rewrite $\sum_j w_{j,0} \big(\sum_k \alpha_{j,k+1} \big) = \sum_j w_{j,0} \alpha_j - \mathcal{E}_{\text{sec}}$ with $\mathcal{E}_{\text{sec}}$ as in (A.2). Collecting terms gives α_t^{fund} equal to holdings alpha plus weight deviation minus trading costs plus $(\mathcal{E}_{\text{comp}} - \mathcal{E}_{\text{sec}})$. Grouping the compounding pieces (and any unobserved friction such as net lending fees) into a single aggregate Residual term then yields the final accounting identity that decomposes the fund alpha into the baseline holdings alpha and an alpha-based formulation of the Kacperczyk et al. (2008) return

gap:

$$\alpha_t^{\text{fund}} \approx \sum_{k=0}^{K-1} \sum_j w_{j,k} \cdot (\alpha_{j,k+1} - \text{TCost}_{j,k+1}) + \text{Residual}$$

$$= \underbrace{\sum_j w_{j,0}\alpha_j}_{\text{(1) holdings alpha}} + \underbrace{\overbrace{\sum_j \sum_{k=0}^{K-1} (w_{j,k} - w_{j,0}) \cdot \alpha_{j,k+1}}^{\text{weight deviation}} - \overbrace{\sum_j \sum_{k=0}^{K-1} w_{j,k}\text{TCost}_{j,k+1}}^{\text{trading costs}} + \text{Residual}}_{\text{(2) Kacperczyk et al. (2008) return gap}}$$

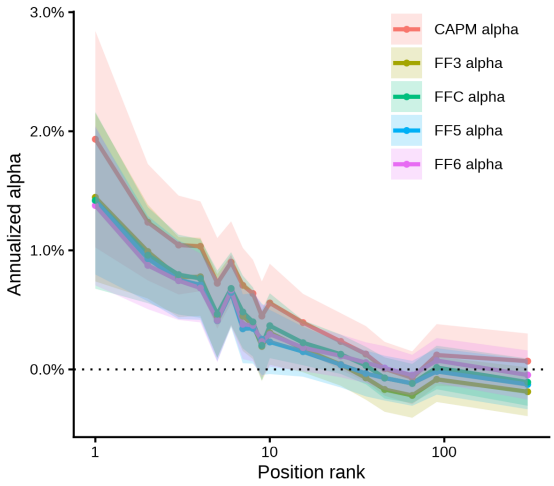
where the time-aggregation part of Residual is ApproximationError $= \mathcal{E}_{\text{comp}} - \mathcal{E}_{\text{sec}}$ as in (A.3). Specifically:

1. **Geometric compounding error** ($\mathcal{E}_{\text{comp}}$): With full-period alpha defined as compounded fund minus compounded benchmark returns, $\mathcal{E}_{\text{comp}} = \mathcal{E}_{\text{comp}}^{\text{fund}} - \mathcal{E}_{\text{comp}}^{\text{bench}}$ as in (A.1). The fund piece is the usual product-minus-sum gap, $\mathcal{E}_{\text{comp}}^{\text{fund}} = \left[\prod_k (1 + R_{k+1}^{\text{fund}}) - 1\right] - \sum_k R_{k+1}^{\text{fund}}$, which a second-order expansion approximates by $-\frac{1}{2}\sum_k (R_{k+1}^{\text{fund}})^2$.
2. **Security-level compounding differences** ($\mathcal{E}_{\text{sec}}$): The holdings-weighted gap between full-period and summed subperiod security alphas, $\mathcal{E}_{\text{sec}} = \sum_j w_{j,0}\left(\alpha_j - \sum_k \alpha_{j,k+1}\right)$ as in (A.2).

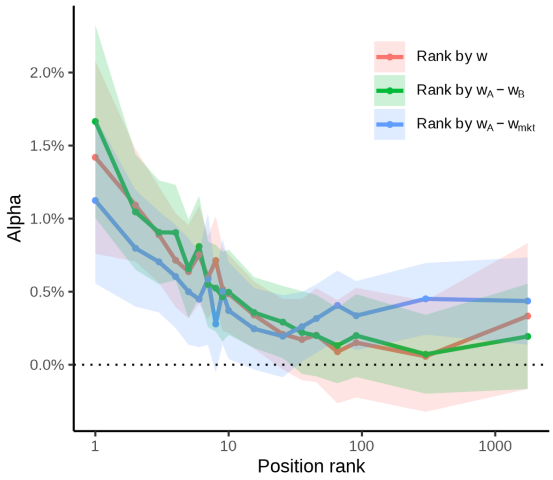
D Holdings alpha DRS

D.1 Robustness of the alpha-by-conviction curve

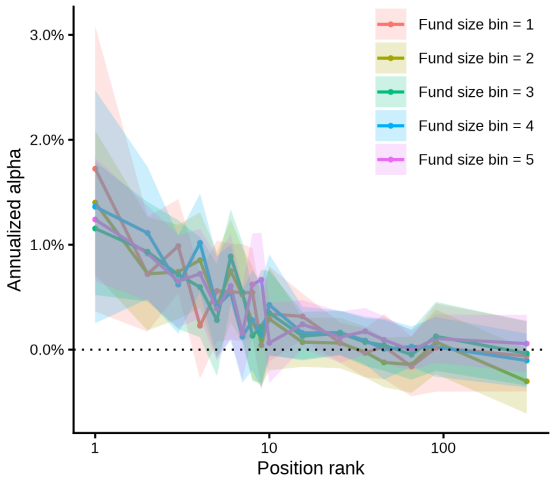
Figure A.2 reports robustness checks on the alpha-by-conviction pattern in Section 4.2. Panel (a) shows that the decline of alpha with position rank is similar under the CAPM, FF3, Carhart (FFC), FF5, and FF6 models. Panel (b) replaces the baseline ranking by active weight relative to the Morningstar style benchmark ($w - w^b$) with rankings by raw portfolio weight (w) and by weight relative to the market portfolio ($w - w^{\text{mkt}}$); the qualitative pattern is unchanged. Panel (c) estimates the FF6 curve separately within fund-size quintiles. The curves are similar across size bins, so the dilution exercise that applies a common $\{\alpha_r\}$ schedule across funds is not an artifact of mixing small and large funds.



(a) Alpha by factor model



(b) Alpha by conviction identification



(c) FF6 alpha by fund size

Figure A.2. Robustness of alpha by position rank

For each fund-quarter, we rank portfolio stocks in descending order of conviction and plot average quarterly alpha by position-rank bin with 95% confidence intervals (log scale on the horizontal axis), as in Figure 1. Panel (a) compares CAPM, FF3, Carhart (FFC), FF5, and FF6 alphas under the baseline ranking by active weight relative to the Morningstar style benchmark. Panel (b) holds the FF6 model fixed and varies the ranking: raw portfolio weight w, active weight relative to the style benchmark $w - w^b$, and weight relative to the market $w - w^{\text{mkt}}$. Panel (c) estimates the FF6 curve separately for funds in each AUM quintile.

D.2 Dilution effect

We follow the discussion in Section 4.2 to estimate the dilution effect on holdings alpha using within fund size variation. For this purpose, we compute hypothetical holdings alphas for

funds:

$$\widehat{\text{HA}}_{i,t} = \sum_r w_{i,r,t} \cdot \alpha_r$$

And we estimate a panel regression of:

$$\widehat{\text{HA}}_{i,t} = a + b \cdot \log_2(A_{i,t-1}) + \tau_i + \tau_t + \epsilon_{i,t}$$

where τ_i and τ_t are fund and time fixed effects, respectively, and $\log_2(A_{i,t-1})$ is the lagged $\log_2$ fund AUM. Because we include fund fixed effect, this isolates within fund size variation. We cluster standard errors by fund and time.

The results are reported in Table A.2. Column (1) shows that, when a fund's size doubles, its hypothetical alpha decreases by about an economically insignificant 0.5 basis points per year, and the effect is statistically insignificant. The subsequent columns interact $\log_2(A)$ with quintile indicators of other fund characteristics to explore heterogeneity of the effect. Throughout all specifications, we find largely consistent results of very minor dilution effects.

Dependent variable: hypothetical fund return (annual, in %)						
Sorted by:		Turnover	Tracking Error	Expense Ratio	Alpha	AUM
	(1)	(2)	(3)	(4)	(5)	(6)
$log_2(A_{i,t-1})$	−0.005	−0.009*	−0.006	0.001	−0.004	−0.013
	(0.004)	(0.005)	(0.005)	(0.005)	(0.008)	(0.010)
$log_2(A_{i,t-1}) \times Bin2$		−0.000	0.001	−0.007*	0.008	−0.010
		(0.003)	(0.005)	(0.004)	(0.008)	(0.013)
$log_2(A_{i,t-1}) \times Bin3$		0.002	−0.002	−0.005	−0.000	0.007
		(0.004)	(0.006)	(0.005)	(0.008)	(0.013)
$log_2(A_{i,t-1}) \times Bin4$		0.004	0.003	−0.004	−0.003	0.003
		(0.005)	(0.006)	(0.006)	(0.009)	(0.013)
$log_2(A_{i,t-1}) \times Bin5$		0.007	0.003	−0.010	−0.008	0.024*
		(0.005)	(0.007)	(0.007)	(0.009)	(0.013)
Time FE	Y	Y	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y	Y	Y
Obs	45,233	45,233	45,233	45,233	45,233	45,233
R^2	0.712	0.716	0.713	0.713	0.713	0.713
Within R^2	0.001	0.014	0.001	0.002	0.002	0.003

Table A.2. Dilution effect: panel regressions with interactions

The dependent variable is the hypothetical fund return (annual, in %), constructed by applying the estimated alpha-by-conviction curve to each fund-quarter's portfolio weights. The key regressor is lagged $\log_2$ fund AUM. Columns (2)–(6) interact $\log_2(A)$ with within-quarter quintile bins of turnover, tracking error, expense ratio, realized alpha, and AUM, respectively. All specifications include fund and time fixed effects. Standard errors are two-way clustered by fund and time.

D.3 Price impact

In the main paper, we use the price impact calibration from Frazzini et al. (2018) (henceforth FIM) as the baseline for f in (5), and Bershova and Rakhlin (2013) (henceforth BR) as a robustness check. Both papers estimate impact from actual parent orders executed by buy-side desks. We describe their settings and then write both curves as $f_{\text{FIM}}(q)$ and $f_{\text{BR}}(q)$ of dollar position $q = Aw$.

Empirical settings. FIM use AQR equity executions over August 1998–June 2016 in 21 developed markets (about 3.5 million parent orders, of which about half are U.S.). Average parent size is 0.9% of the stock's average daily dollar volume V, with a median horizon of 1.7 days. Because our funds are U.S. equity funds, we use their U.S.-only calibration (Table VII, column (10)). BR use AllianceBernstein single-day U.S. equity metaorders from January 2009 to June 2011 (about 12,500 orders after screens). Typical parent size is also about 1% of median daily volume (range 0.1–30%).

Both papers measure peak impact as the signed return from the pre-trade arrival price to the

execution (last-fill) price. The relevant object we seek in (5) is the permanent component f: FIM estimate that about 85% of peak impact remains; BR that about two-thirds remains after post-trade VWAP reversion.

Estimated price impact curves. With slight abuse of notation (relative to Equation (5)), let $f_{\text{FIM}}(q)$ and $f_{\text{BR}}(q)$ denote the two papers' permanent price impact of a dollar position $q = Aw$ in a stock with average daily dollar volume V, or equivalently of a trade of size $\pi = q/V$. Both are in basis points. Both papers find impact concave in π and well described by a square root. FIM's U.S. fit of permanent price impact is:

$$f_{\text{FIM}}(q) = \underbrace{0.85}_{\text{fraction permanent}} \times \underbrace{11.21\sqrt{100\pi}}_{\text{peak price impact}}.$$

BR scale linearly with trailing 30-day daily volatility σ. Their curve is

$$f_{\text{BR}}(q) = 105 \times \sigma \, (2600\pi)^{0.47}.$$

In terms of functional form, the two are similar to each other. FIM started by fitting a more general functional form and then argue that it is sufficiently close to square root. BR estimate the exponent and found 0.47, which is also close to 0.5. FIM also considers volatility, but it enters additively rather than as a multiplier on trade size, so it shifts the level of f and does not change the slope in π. Because DRS comes from how f changes with position size, that additive term does not affect our DRS estimates.

The FIM and BR estimates are comparable to each other. To visualize, Figure A.3 plots the FIM price impact curve in black and the BR curve — evaluated with median daily volatility of stocks in our holdings sample ($\sigma = 1.8\%$) — in blue. When trade size as a fraction of daily volume (π) takes the values of 0.01, 0.05, and 0.1, the estimated price impacts are 9.5, 21.3, and 30.1 bp, respectively, under FIM and 8.7, 18.6, and 25.8 bp, respectively, under BR.

D.4 Liquidity tilt

How much alpha erosion comes from funds tilting toward larger, more liquid stocks as they grow?

Sizing the tilt towards larger stocks. We start by computing the NYSE size decile of each stock j with $\text{SizeDecile}_{j,q} = 1$ or $\text{SizeDecile}_{j,q} = 10$ representing the smallest and largest

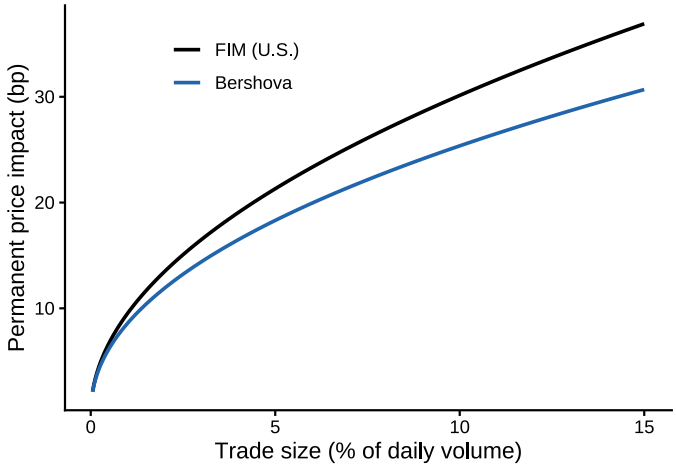


Figure A.3. Permanent price impact as a function of trade size

This figure plots permanent price impact (in basis points) against trade size as a fraction of daily volume. The black line is the U.S. calibration of Frazzini et al. (2018). The blue line is Bershova and Rakhlin (2013), evaluated at the median daily volatility of stocks in our holdings sample ($\sigma = 1.8\%$).

stocks, respectively. We then compute the value-weighted average size decile for each fund i in quarter q:

$$\text{SizeDecile}_{i,q}^{\text{fund}} = \sum_j w_{i,j,q} \cdot \text{SizeDecile}_{j,q}$$

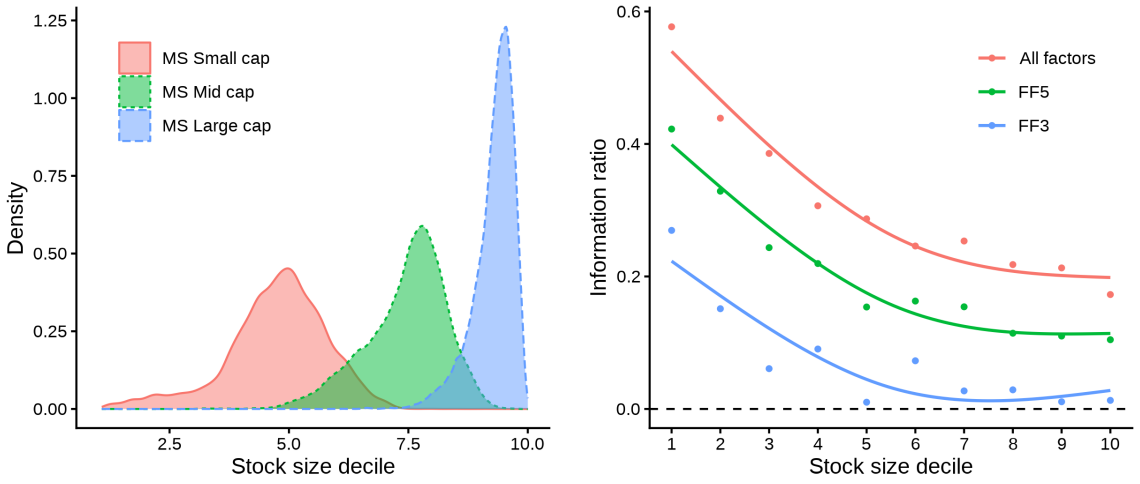
where $w_{i,j,q}$ is the weight of stock j in fund i's portfolio in quarter q. Panel (a) of Figure A.4 shows the distribution of $\text{SizeDecile}_{i,q}^{\text{fund}}$ by Morningstar fund size style. The median size decile is roughly 5 for small-cap funds, 8 for mid-cap funds, and 9.5 for large-cap funds.

We now estimate how much this metric changes with fund size. We estimate a panel regression of:

$$\text{SizeDecile}_{i,q}^{\text{fund}} = a + b \cdot \log_2(A_{i,q-1}) + \tau_i + \tau_q + \epsilon_{i,q}$$

where the dependent variable is the portfolio value-weighted average holdings size decile of fund i in quarter q, and where τ_i and τ_q are fund and time fixed effects, respectively. The coefficient b estimates the change in holding size decile when fund size doubles. We cluster standard errors by quarter and fund.

The results are reported in Table A.3. Column (1) does not contain fund fixed effects and thus also makes use of across-fund variation, and finds that fund size doubling is associated with a 0.167 increase in holding size decile. Column (2) contains fund fixed effects to only



(a) (b)

Figure A.4. The effect of holdings liquidity on achievable fund returns

Panel (a) assigns each stock a NYSE size decile (1 = smallest, 10 = largest) and computes each fund-quarter's portfolio value-weighted average holdings size decile. It plots the kernel density of this measure separately for Morningstar small-, mid-, and large-cap funds. Small-cap funds center near decile 5, mid-cap funds near decile 8, and large-cap funds are tightly concentrated near decile 10. Panel (b) plots achievable information ratios of long–short characteristic portfolios formed within each NYSE size decile (separately for FF3, FF5, and all characteristics), with fitted splines.

make use of within-fund variation — which is appropriate for estimating fund-level DRS — and finds that the effect declines to 0.046. Columns (3) and (4) interact the independent variable with Morningstar capitalization and growth styles. The results indicate that the effect is stronger for small-cap and growth funds, but those coefficients are all lower than 0.1. To put this into perspective, consider a typical small-cap fund which has a holding size decile of around 5. When its size doubles, it will slightly change its portfolio so that the metric only rises to below 5.1.

As robustness, columns (5) through (8) of Table A.3 compute fund-level holdings size decile using the absolute value of active fund positions as weights. Specifically, we compute:

$$\text{SizeDecile}_{i,q}^{\text{fund}} = \frac{\sum_j |w_{i,j,q} - w_{i,j,q}^b| \cdot \text{SizeDecile}_{j,q}}{\sum_j |w_{i,j,q} - w_{i,j,q}^b|}$$

where $w_{i,j,q}^b$ is the benchmark weight defined by the aggregate portfolio of funds in the corresponding 3×3 Morningstar size-style group. This measure isolates the liquidity of each fund's active deviation from benchmark. As shown in the Table, the results are broadly similar with even lower magnitudes. Overall, we find that funds only mildly tilt towards

large, liquid stocks when their sizes increase.

	Dependent variable: $\text{SizeDecile}_{i,q}^{\text{fund}}$							
Measure based on:	Holdings				Active holdings			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log_2(A_{i,q-1})$	0.167***	0.046***			0.137***	0.020***		
	(0.016)	(0.005)			(0.014)	(0.004)		
$\log_2(A_{i,q-1}) \times$ Small			0.090***				0.063***	
			(0.012)				(0.010)	
$\log_2(A_{i,q-1}) \times$ Mid			0.073***				0.029***	
			(0.009)				(0.006)	
$\log_2(A_{i,q-1}) \times$ Large			0.005				−0.010	
			(0.008)				(0.006)	
$\log_2(A_{i,q-1}) \times$ Value				0.030***				0.006
				(0.009)				(0.008)
$\log_2(A_{i,q-1}) \times$ Blend				0.050***				0.023***
				(0.010)				(0.006)
$\log_2(A_{i,q-1}) \times$ Growth				0.052***				0.027***
				(0.007)				(0.005)
Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y
Fund FE	N	Y	Y	Y	N	Y	Y	Y
Obs	67,381	67,381	67,381	67,381	67,381	67,381	67,381	67,381
R^2	0.055	0.972	0.972	0.972	0.053	0.983	0.984	0.983
Within R^2	0.041	0.017	0.031	0.018	0.034	0.007	0.025	0.008

Table A.3. Holdings liquidity by fund size

We classify stocks into NYSE size deciles (1 to 10) and compute the holding value-weighted size decile by fund-quarter. Columns (1)–(4) use actual portfolio weights; columns (5)–(8) use absolute deviations of portfolio weights from style-benchmark weights. The table reports panel regressions of this measure on lagged fund size ($A_{i,q-1}$) in base-2 logs. Columns (1) and (5) include only quarter fixed effects; the remaining columns also include fund fixed effects. Columns (3)–(4) and (7)–(8) interact lagged fund size with Morningstar style categories. Standard errors are clustered by quarter and fund. The last row reports within R^2. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Sizing the impact on fund returns. How much of a drop in fund returns can this explain? To get a sense of how achievable returns depend on stock size, we investigate this using well-known characteristics-based stock factors. Over 1970–2022, within each NYSE size decile we sort stocks into quintiles by a characteristic, value-weight within bins, and form long–short returns. Thus, for each characteristic — book-to-market, market beta, gross profitability, asset growth, accruals, net operating assets, momentum, industry momentum, intermediate momentum, seasonal momentum, short- and intermediate-term reversal, one- and five-year issuance, and realized volatility — we obtain a size-conditional long–short factor. We use the variation of these long–short factor returns across size deciles to gauge the degree to which achievable returns decline with stock size.

Panel (b) of Figure A.4 plots the average information ratios against stock size decile with fitted splines. The blue, green, and red lines represent averages over characteristics in the Fama-French three-factor model, five-factor model, and all characteristics, respectively. The figure is strongly consistent with the idea that achievable returns decline with stock size. For instance, for the "all factors" group, the average information ratio is close to 0.6 for the bottom stock size decile but declines to around 0.2 for the top size decile. On average, this means that each one-unit increase in the size decile reduces the information ratio by around $(0.6 - 0.2)/(10 - 1) \approx 0.044$. Because the typical tracking error of mutual funds is around 6% (annualizing from the standard deviation of benchmark-adjusted return in Table 1), a one-unit increase in holdings size decile implies a decline in annual fund alpha of around $6\% \times 0.044 = 0.264\%$.

We now put these estimates together to quantify the DRS effect from liquidity tilts. Table A.3 finds that, when a fund's size doubles, the change in holdings size decile is on average 0.046. If the fund's active strategy has similar properties to the stock factors discussed above, this can lead to a decline of annual fund return of $0.046 \times 0.264\% \approx 0.01\%$. For the small-cap funds whose liquidity tilt is stronger, the effect is slightly larger, at around $0.090 \times 0.264\% \approx 0.02\%$. This analysis indicates that the liquidity tilt has limited contribution to DRS effects.

E Return gap DRS

This appendix documents two facts that discipline the return-gap mechanisms in Section 5.2.

Return-gap volatility and turnover. To the extent that the weight-deviation component of the return gap is driven by trading, funds that trade more actively should exhibit more variable return gaps. Figure A.5 sorts funds into within-month turnover deciles and, within each decile, computes the pooled standard deviation of the monthly return gap across fund-months. It then plots that standard deviation against the decile-mean turnover ratio. Dispersion rises monotonically: the top-turnover decile has a return-gap standard deviation about 2.7 times that of the bottom decile. In robustness checks, we have also proxied for pure intraquarter trading by subtracting 13F holdings-implied turnover from overall turnover, and find similar results.

Turnover and fund size. If larger funds are less nimble in taking advantage of fast moving opportunities, we should observe that turnover declines with fund size. Table A.4 reports panel regressions of the CRSP turnover ratio on lagged $\log_2$ AUM, with fund and month fixed

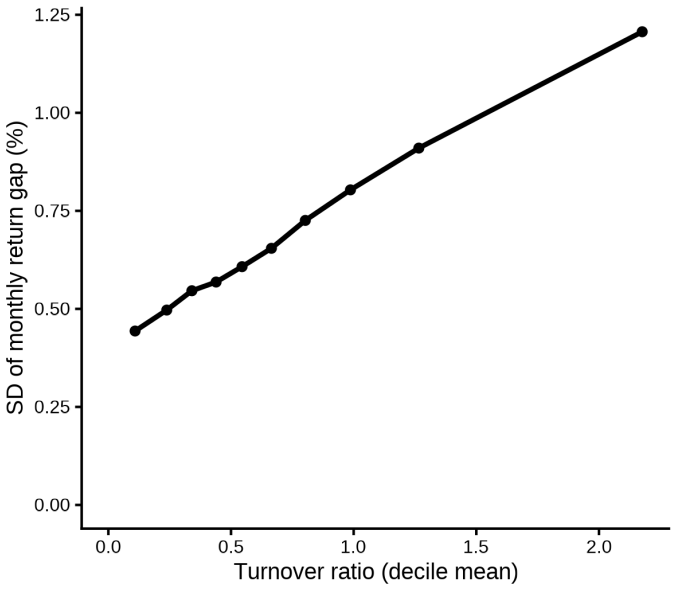


Figure A.5. Return-gap volatility by turnover

Each month, we sort diversified active U.S. equity funds into deciles by CRSP turnover ratio. For each decile, we pool all fund-months and compute the standard deviation of the monthly return gap (in percent), and plot it against the mean turnover ratio in that decile. Return gap is gross fund return minus holdings-implied return, following Kacperczyk et al. (2008).

effects and standard errors clustered by fund and month. Column (1) shows that when a fund's size doubles, its turnover falls by about four percentage points on average. Columns (2) and (3) allow separate slopes by Morningstar capitalization style and by value/blend/growth style; the coefficients remain tightly clustered around -0.04 and are statistically significant throughout. This decline can be compared to the average and median turnovers of 76% and 59%, respectively (Table 1).

Dependent variable: CRSP turnover ratio			
	(1)	(2)	(3)
$\log_2(A_{i,t-1})$	-0.040^{***}		
	(0.005)		
$\log_2(A_{i,t-1}) \times$ Small		-0.041^{***}	
		(0.006)	
$\log_2(A_{i,t-1}) \times$ Mid		-0.036^{***}	
		(0.005)	
$\log_2(A_{i,t-1}) \times$ Large		-0.041^{***}	
		(0.005)	
$\log_2(A_{i,t-1}) \times$ Value			-0.043^{***}
			(0.005)
$\log_2(A_{i,t-1}) \times$ Blend			-0.041^{***}
			(0.005)
$\log_2(A_{i,t-1}) \times$ Growth			-0.038^{***}
			(0.005)
Month FE	Y	Y	Y
Fund FE	Y	Y	Y
Obs	281,651	281,651	281,651
R^2	0.594	0.594	0.594
Within R^2	0.010	0.010	0.010

Table A.4. Turnover and fund size

The table presents panel regressions of the CRSP turnover ratio on lagged fund size in base-2 logs. Column (1) is the full-sample slope. Column (2) estimates separate slopes by Morningstar capitalization style (small, mid, large). Column (3) estimates separate slopes by Morningstar growth style (value, blend, growth). All specifications include month and fund fixed effects. Standard errors are clustered by month and fund. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

F Additional empirical results

F.1 Industry-level and peer DRS

Our paper focuses on the fund-level DRS effects proposed by BG. In addition, Pástor and Stambaugh (2012) also propose that there are industry-level DRS effects: mutual funds generate higher alphas when the entire active mutual fund sector is smaller. Pástor et al. (2015) provides empirical evidence for industry-level DRS effects.

The existence of industry-level DRS does not impact our main conclusion about the quantitative importance of BG. For instance, consider the analysis on performance persistence in Section 6 where we show that fund-level DRS cannot quantitatively explain much of the low persistence of realized alphas. Note that the focus is on the *relative* performance of funds: those with higher alpha in one period generally do not continue to have higher alphas in the subsequent period. Industry-level DRS effects, to a first-order approximation, impact all active funds equally and thus do not impact the *relative* performance of funds.

While DRS at the level of the entire mutual fund sector does not impact our main conclusion, one may wonder whether peer DRS effects do. Intuitively, each fund may be matched to a set of other peer funds which may seek to generate alpha in stocks with similar characteristics. For instance, two funds that both focus on finding opportunities within small-cap growth stocks are close peers. If the two funds tend to have highly overlapping alpha ideas, fund size changes in one of them may plausibly affect the others' ability to generate alpha. Further, if fund flows among peer funds tend to be correlated, then such peer DRS may play a further dampening effect on the persistence of fund performance above and beyond that of the fund-level effects.

To investigate this possibility, we follow Hoberg et al. (2018) to measure pairwise similarity between funds using the distance of characteristics of the stocks held. We briefly explain the methodology and refer the reader to their original paper for details. We first turn stock size, book/market, and momentum characteristics into z-scores using the NYSE subsample. Then, those characteristics z-scores are summarized at the fund-level using portfolio-weighted averages. Finally, pairwise distances of funds are computed by quarter as the Euclidean distance between fund-level characteristics.

For each fund, we follow Hoberg et al. (2018) to classify all other funds with distance — measured using the latest quarterly holdings — less than d_1^* as "group 1" peers, where d_1^* is set such that the average number of peers for an average fund is 8.858%. They chose this number so that the size of peer groups can be compared to that of Lipper categories. To also include funds that are slightly further away, we also define group 2 peers as those with distance less than d_2^*, where d_2^* is defined so that the average number of funds in group 2 matches that of group 1. Group 3 is defined accordingly. Finally, for each fund, we add up the AUM of funds in each group, and we enrich the return gap regressions of (8) by including those peer group sizes as well:[8]

$$\text{ReturnGap}_{i,t} = -\,b \cdot \log_2(A_{i,t-1}) - \sum_{g\in\{1,2,3\}} b_g^{\text{peer}} \cdot \log_2(A_{i,t-1}^{\text{peer }g}) + \tau_t + \tau_i + \epsilon_{i,t} \qquad \text{(A.8)}$$

As in (8), we include month and fund fixed effects and cluster standard errors by month and fund. Here b and b_g^{peer} are DRS magnitudes, so the coefficients on the $\log_2$ size terms in the table equal $-b$ and $-b_g^{\text{peer}}$.

If there are peer DRS effects, we expect the b_g^{peer} magnitudes to be positive. Further, eco-

[8] We focus on the effect through return gap because the analysis in Section 5 shows that this is the dominant part of DRS effects.

nomic logic suggests that the effect from more closely connected peers should be larger, and thus we expect $b_1^{\text{peer}} > b_2^{\text{peer}} > b_3^{\text{peer}}$.

Dependent variable: annualized return gap ($\text{ReturnGap}_{i,t}$, in %)							
Distance measure:	N/A	Hoberg et al. (2018)			Wahal and Wang (2011)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\log_2(A_{i,t-1})$	-0.242^{***}	-0.244^{***}	-0.245^{***}	-0.244^{***}	-0.242^{***}	-0.242^{***}	-0.242^{***}
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
$\log_2(A_{i,t-1}^{\text{peer group 1}})$		0.008	0.060^{**}	0.051^{*}	-0.001	0.008	0.013
		(0.009)	(0.027)	(0.028)	(0.012)	(0.011)	(0.011)
$\log_2(A_{i,t-1}^{\text{peer group 2}})$			-0.054^{**}	-0.077^{**}		-0.015	-0.004
			(0.026)	(0.033)		(0.012)	(0.012)
$\log_2(A_{i,t-1}^{\text{peer group 3}})$				0.033			-0.018
				(0.031)			(0.013)
Month FE	Y	Y	Y	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y	Y	Y	Y
Obs	281,651	274,499	274,499	274,499	277,750	277,750	277,750
R^2	0.074	0.075	0.075	0.075	0.075	0.075	0.075

Table A.5. Inferring peer-level DRS effects

This table presents panel regressions of annualized fund return gap (in percent) on lagged own fund size and the lagged sizes of peer funds, with size transformed to base-2 logarithm so that coefficients measure the effect on annual return when size doubles. Column (1) includes only own size. Columns (2)–(4) add the average lagged size of increasingly broad peer groups defined by holdings-based style distance following Hoberg et al. (2018); columns (5)–(7) repeat the exercise using the holdings-overlap distance of Wahal and Wang (2011). Peer group 1 is the nearest peer set, and groups 2 and 3 successively expand to more distant peers. All regressions include month and fund fixed effects. Standard errors are clustered by month and fund. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

The results are reported in Table A.5. For comparison, column (1) omits peer group sizes and recovers the familiar own-fund DRS estimate: doubling own size reduces annualized return gap by about 0.24%. Columns (2)–(4) successively add the three peer-group size controls. Across these specifications, the own-size coefficient is essentially unchanged. Column (2) does not show any evidence of peer-group DRS from the most closely aligned peers. We do find some statistically significant evidence for peer-level DRS for group 2 but the effect is much smaller than the own-group effect; there is no evidence for group 3. Overall, the evidence for economically meaningful peer DRS is limited.

One may be concerned that measuring distances between funds using stock characteristics is not granular enough. Inspired by Wahal and Wang (2011), we also compute a holdings-based pairwise distance between funds. Specifically, for funds i and k in quarter t, we define

distance as:

$$\text{dist}_{i,k,t} = \frac{\sum_{\text{stock } j}(w_{i,j,t} - w_{k,j,t})^2}{\sqrt{\left(\sum_{\text{stock } j} w_{i,j,t}^2\right) \cdot \left(\sum_{\text{stock } j} w_{k,j,t}^2\right)}}$$

where $w_{i,j,t}$ is the portfolio weight of fund i in stock j. We use the same procedure described above to define peer groups 1, 2, and 3 for each fund and rerun the regressions. The results, reported in columns (5) through (7) in Table A.5, also do not indicate any clear peer DRS effects. Overall, we do not find clear evidence of peer DRS effects.