JOIM
www.joim.com
JOURNAL OF INVESTMENT MANAGEMENT, Vol. 19, No. 2, (2021), pp. 4–28

WHAT HAPPENS WITH MORE FUNDS THAN STOCKS? ANALYSIS OF CROWDING IN STYLE FACTORS AND INDIVIDUAL EQUITIES

Ananth Madhavan, Aleksander Sobczyk*, and Andrew Ang*,†*

The proliferation of funds juxtaposed against the decline in individual stock listing since the mid-1990s raises questions about crowding in individual stocks or style factors. We examine these issues by characterizing the common components of funds from 2007 through 2018. A key difference from the previous literature on common factors in fund returns is that we explicitly look at fund holdings over time for all US-listed equity active mutual funds and exchange-traded funds and contrast their differences. We also explore the implications of this proliferation in funds for the pricing of individual securities and funds.

1 Introduction

In the US today, the number of mutual funds and exchange-traded funds focused on domestic equities now exceeds the number of US-listed stocks.[1] The proliferation of funds raises important economic questions: What are the most common positions held by funds, and do they differ by fund types like mutual funds and exchange-traded funds (ETFs)? Do certain funds hold concentrated portfolios in individual stocks or factor exposures, and is there evidence of increased crowding over time? Are funds becoming more or less diverse in their investment strategies?

This paper examines these questions by characterizing the common components of funds, in individual stocks and exposure to style factors, over the period January 1, 2007 to December 31, 2018. Our dataset consists of the holdings of all US-listed, US-equity focused mutual funds and ETFs. A key difference from an extensive previous literature on common factors in fund returns (see, for example, Connor and Korajczyk, 1988) and hedge fund replication (see, for example, Hasanhodzic and Lo, 2007) is that we explicitly look at individual fund holdings over time. Time-series data on holdings provide the most granular look at similarities in positions across funds, and a further mapping of those holdings to style factors

*BlackRock Inc., 400 Howard Street, San Francisco, CA 94105, USA.
†The views expressed here are those of the authors alone and not of BlackRock, Inc. We thank Ronald Kahn for his many helpful comments. This material is not intended to be relied upon as a forecast, research or investment advice, and is not a recommendation, offer or solicitation to buy or sell any securities or to adopt any investment strategy.

allows us to assess the extent to which popular style factors have seen increased crowding.

An important element of our analysis is to distinguish crowding among active managers and in ETFs. This distinction is interesting since the channels for concentration in styles or stocks may differ by investment vehicle. For active funds, managers using similar data and strategies, might gravitate to the same sorts of positions, either in individual names or in factors as described by Khandani and Lo (2011) in the "Quant Quake" of 2007. We simply do not know, if in recent years, whether manager strategies are becoming more similar as factor and ESG investing gain in popularity or more diverse as managers seek differentiation in data and styles. Indeed, studies of crowding focus on active funds like mutual funds (e.g., Coval and Stafford, 2007) or hedge funds (e.g., Brown *et al*., 2019), but because ETFs are open-ended, these vehicles are also subject to crowding as well.

In ETFs, possible crowding may arise for different reasons than active funds. ETFs are often thought of as passive vehicles, since the vast majority of funds seek to track a given index, but many active investors use ETFs as building blocks to gain exposure to sectors, industries, or factors. This raises the possibility that crowding could arise because exposures in open-end funds are determined by investor flows that may exhibit commonality. For example, if investors chase stock returns there could be increased flows into funds exposed to momentum either explicitly (through momentum factor funds) or implicitly through funds oriented to sectors/strategies/stocks that have seen high past returns, even these funds do not deem themselves as factor funds.

To our knowledge, this is the first paper to analyze questions of crowding across ETFs using holdings data. The universe of ETFs is also interesting because some commentators have expressed concerns that investors will express their views using funds rather than stocks, causing liquidity in individual names to dry up. While we believe such concerns are not valid, it is nonetheless interesting to address the question of whether individual securities can be valued solely using fund prices. We show that this is, in fact, the case but this equivalence has held only relatively recently.

To measure crowding, we create a high-dimensional weighting matrix of fund holdings with as many rows as there are US-listed stocks and as many columns as there are long-only funds. We do this separately for US-listed and US-equity active mutual funds and ETFs every quarter. Each column represents a specific fund; reading down the rows for a given column, we obtain that fund's portfolio weights (a fraction between 0 and 1) across the universe of stocks so that the column sums add to 1.

The matrix of fund positions (either stock weights or dollars) has noteworthy features. First, it is "sparse" because many managers—particularly those following fundamental investment strategies based on individual company analysis—hold concentrated positions of a few dozen stocks, so that many of the entries in a given row in the matrix are exactly zero. Second, some of the columns of the holding matrix may be very similar because many funds share the same benchmark (e.g., the S&P 500 index), and some columns may be linear combinations of others (e.g., funds whose benchmark is the Russell 3000, which in turn comprises the Russell 1000 and Russell 2000).[2] Third, the matrix can easily be permuted allowing us to analyze exposure to a linear multi-factor model of equity returns to understand how funds are alike, not only in holdings, but also in their exposures to a much smaller set of rewarded factors.

We define crowding in terms of a very specific dimensionality question about this matrix, i.e., the extent to which information in the fund weight matrix can be summarized efficiently by a subset of funds or combinations of stocks.[3] Applying this analogy to the weighting matrix for holdings (by quarter from 2000 to 2017) yields the "canonical" investment themes over time, and thus this serves as a natural metric for crowding. These are of interest because of a growing literature on scale and fees in asset management and the equilibrium economics of active and index funds (Pastor *et al*., 2015; Stambaugh, 2019). It is worth noting that, unlike ETFs where redemptions can be made in-kind (see Madhavan, 2016), mutual fund net redemptions result in flows in the underlying stocks. Hence, it is important for analyses of crowding to distinguish between these types of open-end funds. Thus, for these and other reasons, we break out analysis with mutual funds and ETFs separately in our analysis.

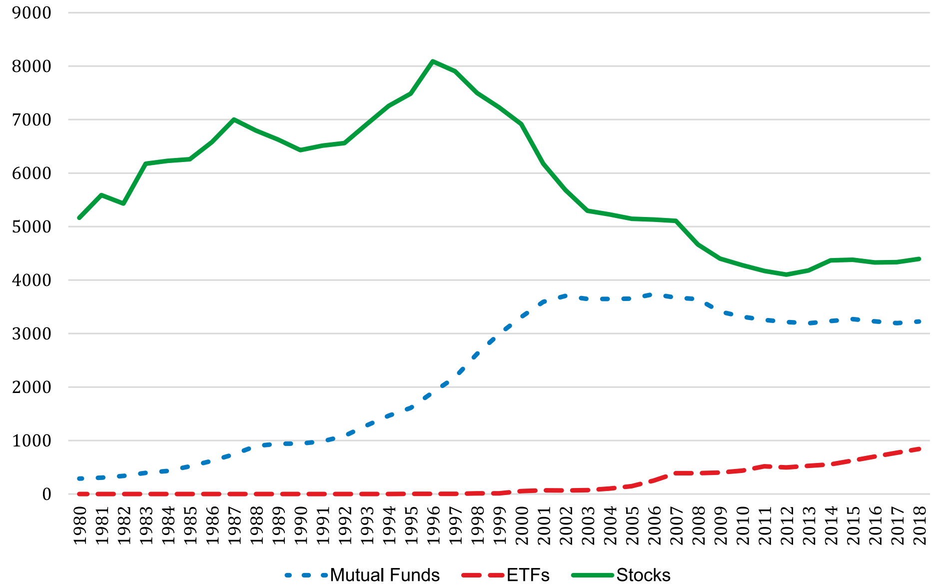
We show that despite the large number of funds, fewer than a dozen fund canonical portfolios ("archetypes") well describe the whole universe of funds. The commonality of active mutual funds has remained approximately constant in the period 2007 to 2018 so that in the aggregate, there is no evidence of increased crowding. However, within style box categorizations, we find evidence of increased crowding over the last decade among growth-oriented mutual funds. Factor crowding in common style factor exposures has remained relatively constant. Turning to the set of all ETFs, we find decreases in common ETF positions over the period 2007–2018. Furthermore, mutual fund positions are more alike than ETFs consistent with the proliferation of ETFs tracking many disparate indexes. Fears that the proliferation of funds relative to the universe of stocks has increased crowding are not supported by the empirical evidence. The analysis also helps us understand the characteristics of portfolios that are common across funds, and their attributes.

Our work relates primarily to a large literature measuring factor loadings of funds. These have typically used time-series regressions with unconstrained Fama and French (1993) value, size, and other style factors or constrained regressions following Sharpe (1992). Daniel and Titman (1997) use characteristics of individual stock holdings as opposed to estimating factor loadings. Authors like Daniel *et al*. (1997), Wermers (2000, 2012), Jiang *et al*. (2007), and Brown *et al.* (2009) use either or both of these techniques and document that factor exposures of funds do vary over time. These papers, unlike our approach, do not directly estimate crowding measures of funds. A more recent literature tries to directly estimate crowding of style factors. These authors, like Bonne *et al.* (2018), Baltas (2019), and Marks and Shang (2019) generally use time series of returns to compute covariances, augmented with factors, characteristics, and liquidity measures. In contrast, our approach estimates commonality with cross-sectional stock weights of all funds. We also give insights into the stocks held by typical funds and characterize the most common types of funds.

2 Funds and Stocks

Before turning to our data sources and empirical results, some institutional context regarding trends in US stocks and funds over the past four decades provides useful context. Exhibit 1 shows the total number of US-listed stocks and US-listed, US equity-focused mutual funds and equity ETFs from 1980 to 2018.[4] At the end of 2018, there are 8,078 US-listed mutual funds of which 4,753 are purely equity-focused and 1,988 US-listed ETFs, of which 1,510 are purely equity-focused. By contrast, the number of US exchange-listed companies at this time is 4,397. The number of US-listed equity funds (focused

Exhibit 1 Number of US-listed, US-focused equity mutual funds, equity ETFs, and stocks: 1980–2018.



Data on the number of US stocks is drawn from the World Bank (2019); data on funds is sourced from the Investment Company Institute Fact Book (2019). The numbers for ETFs and the mutual funds include only pure US-focused equity funds that are domiciled in the US, and exclude bond funds, money market funds, commodity funds, international funds, and hybrid funds. Mutual funds or ETFs that invest in other funds exclusively and closed-end funds/unit trusts are excluded.

on the US) first exceeds the number of US-listed stocks in 2007, and the gap continues to widen since that time. In our later empirical work, we will use a subset of US-focused mutual funds and ETFs that are largely representative of this broad universe.

We observe in Exhibit 1 changes in investor preference for the type of fund wrapper; the growth in ETFs is particularly noteworthy. The first US-listed ETF was incepted in 1993, and by 2018 the number of US-listed equity ETFs grows to 1,510, with rapid growth after the mid-2000s.[5] As of December 31, 2018, there are $4.7 trillion of ETF assets worldwide representing approximately 5% of the market capitalization of the global market, up from zero in 1993. Even within the United States—the largest ETF market—just 9% of the total assets invested in US equities are in US-listed equity ETFs.[6] Interestingly, the number of mutual funds plateaus *before* the rapid increase in ETFs during the mid-2000s, so it is not explosion in ETF listings *per se* that causes the number of mutual funds to stagnate.

What explains the long-term trends in US stock listing? Authors like Doidge *et al.* (2017) argue that regulation has increased the cost of going public in the US, with underwriting and registration costs estimated as high as 14% of the amounts raised. This is a significant hurdle for smaller companies, and indeed, much of the reduction in US-listed companies in recent years is in micro-cap stocks. At the same time, the growth of private equity and alternatives investing may provide a lower cost of capital to firms that elect not to go public, while the rise of matching services allows private company equity holders to obtain liquidity. Finally, mergers and acquisitions activity in the US in recent decades also play a role in reducing the number of companies.

What explains the growth in the number and types of funds? In a complete Arrow–Debreu world, a fund is simply a repackaging of existing securities and there is no unique information in the fund versus the underlying stocks, and the number of funds, and choice of wrapper, is indeterminate. In practice, fund management companies have a comparative advantage in selecting and trading portfolios of stocks more efficiently than individual investors that derives from economies of scale and, increasingly, technological advantages in collecting and processing large volumes of data. Furthermore, individuals may also have preferences for certain styles of investing or are reached by different distribution channels that are best suited to different types of funds, so it is perhaps not surprising that there are now more funds than stocks.

The economics of asset management also favors creating new funds. For active funds, managers may incubate several different strategies before bringing these to market, given that investor flows to differentiated fund features.[7] New index strategies are often differentiated as well—across sectors, regions, or systematic investment styles. Furthermore, there may be a winner-takes-all phenomenon where the first fund to launch in a category that is new to market attracts attention and the majority of flows.

3 Numerical Example

In this section, we present an example that illustrates our approach to summarizing the information content of fund holdings, with the technical details contained in the Appendix. We work with quarterly holding matrices $\boldsymbol{W}$ where each row is a stock ($i = 1, \ldots, n$) and each column ($j = 1, \ldots, p$) represents a fund. Each cell (i, j) in $\boldsymbol{W}$ represents the weight (0,1) of stock i in fund j, so that the column sums are exactly 1.[8] For expositional simplicity, we assume that there are six funds ($p = 6$ columns) and the universe of stocks is five ($n = 5$ rows). We write the weights, in up to five stocks, held by the six funds in a 6×5 matrix, $\boldsymbol{W}$:

	Fund 1	Fund 2	Fund 3	Fund 4	Fund 5	Fund 6
Stock 1	0.5	0.2	0.4	0.2	0.2	0.5
Stock 2	0.5	0.7	0.0	0.1	0.2	0.2
Stock 3	0.0	0.0	0.0	0.0	0.2	0.1
Stock 4	0.0	0.1	0.6	0.7	0.2	0.1
Stock 5	0.0	0.0	0.0	0.0	0.2	0.1

The matrix $\boldsymbol{W}$ has several characteristics that are observed in our actual data. First, some funds hold concentrated positions, like Funds 1 and 3 which hold only two stocks. In data, high conviction fundamental managers typically hold relatively few stock positions. This causes $\boldsymbol{W}$ to have many entries of exactly zero, or in technical terms it is a "sparse" matrix. Other funds like Funds 5 and 6 hold relatively large numbers of stocks, similar to market capitalization index, smart beta, or quantitative alpha-seeking funds in data.

3.1 First approximation

We can approximate the weighting matrix, $\boldsymbol{W}$, by the following product, termed a singular value decomposition (SVD):

$$\tilde{W}^{(\mathbf{1})} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{1}$$

where

$$\tilde{W}^{(\mathbf{1})} = \begin{bmatrix} 0.633 \\ 0.541 \\ 0.073 \\ 0.544 \\ 0.073 \end{bmatrix} [1.30] \times [0.451 \quad 0.430 \quad 0.446 \quad 0.432 \quad 0.287 \quad 0.380]$$

The superscript "(1)" in $\tilde{W}^{(\mathbf{1})}$ denotes that this is the first approximation for $\boldsymbol{W}$. The representation is low dimensional because we are trying to capture a matrix with 30 different entries using just

12(= 5 + 1 + 6) numbers. Multiplying this out, we have

$$\tilde{W}^{(1)} = \begin{bmatrix} 0.371 & 0.354 & 0.367 & 0.356 & 0.236 & 0.312 \\ 0.317 & 0.302 & 0.313 & 0.304 & 0.202 & 0.267 \\ 0.043 & 0.041 & 0.043 & 0.041 & 0.027 & 0.036 \\ 0.319 & 0.305 & 0.316 & 0.306 & 0.203 & 0.269 \\ 0.043 & 0.041 & 0.043 & 0.041 & 0.027 & 0.036 \end{bmatrix}$$

Compare the stock weights held by Fund 1 in $\boldsymbol{W}$, which is given by 0.50 on Stock 1 and 0.50 on Stock 2, with the approximation given by the first column of $\tilde{W}^{(1)}$:

	Stock 1	Stock 2	Stock 3	Stock 4	Stock 5
Fund 1 $\tilde{W}^{(1)}$ Weights	0.371	0.317	0.043	0.319	0.043
Actual Fund 1 Weights	0.5	0.5	0	0	0

This first pass captures the large weights on Stock 1 and Stock 2 (with weights of 0.371 and 0.317, respectively), but has some approximation error, especially with the relatively high weight on Stock 4 (which is supposed to have zero weight).

The matrix $\boldsymbol{U} = [0.633\ 0.541\ 0.073\ 0.544\ 0.073]^T$ can be interpreted as the first combination of *stocks* that play the largest role in explaining joint positions across all the funds. It is called the *left singular matrix* which is a unitless vector in the sense that the variance of the positions sum to 1, that is, $\boldsymbol{U}^T\boldsymbol{U} = 1$. If we compute the average stock weights across the funds, and rescale them such that they sum to one, we obtain $[0.634\ 0.539\ 0.95\ 0.539\ 0.095]^T$—which is very close to $\boldsymbol{U}$. In fact, when we compute this approximation in data, the first approximation does correspond closely to an average fund weight of stocks.

The matrix $\boldsymbol{V}^T = [0.451\ 0.430\ 0.446\ 0.432\ 0.289\ 0.380]^T$ can be interpreted as the first combination of *funds* that bring us close to matching the typical weights of all funds in $\boldsymbol{W}$. The matrix $\boldsymbol{V}$ is called the *right singular matrix*. In our case, the combination of funds in $\boldsymbol{V}$ is close to an average because the entries in $\boldsymbol{V}$ are approximately the same. Like the matrix $\boldsymbol{U}$, the square of the entries sum to 1, that is, $\boldsymbol{V}^T\boldsymbol{V} = 1$. Below, we also obtain this result that to a first-order approximation in data: the first and most important characteristic is the stock weights of a typical fund—a simple estimate of this is to compute the average stock weights held by all funds at a point in time.

The number 1.30 in the $\boldsymbol{S}$ matrix (which is scalar in this first iteration) is called a *singular value*. It is a scaling variable, which we apply to the typical stock weights in $\boldsymbol{U}$ held by the typical fund, approximated by the fund weights in $\boldsymbol{V}$, to bring us to match the magnitude of the weights in the data matrix $\boldsymbol{W}$. We scale because the combinations in $\boldsymbol{U}$ and $\boldsymbol{V}$ are set to have unit length. We can consider rotating, or combining, stocks (the weights in $\boldsymbol{U}$) or rotating, or combining, funds (the weights in $\boldsymbol{V}$), and then scale them by $\boldsymbol{S}$ to match the data.

3.2 Second approximation

Now consider the following second approximation, $\tilde{W}^{(2)}$, of the data matrix $\boldsymbol{W}$ with new $\boldsymbol{U}$, $\boldsymbol{S}$, and $\boldsymbol{V}$ matrixes:

$$\tilde{W}^{(\mathbf{2})} = \boldsymbol{USV}^T,$$

where

$$\tilde{W}^{(\mathbf{2})} = \begin{bmatrix} 0.633 & -0.090 \\ 0.541 & -0.650 \\ 0.073 & -0.016 \\ 0.544 & 0.754 \\ 0.073 & -0.016 \end{bmatrix} \begin{bmatrix} 1.30 & 0 \\ 0 & 0.83 \end{bmatrix} \times \begin{bmatrix} 0.451 & 0.430 & 0.446 & 0.432 & 0.287 & 0.380 \\ -0.450 & -0.483 & 0.506 & 0.541 & -0.004 & -0.125 \end{bmatrix}.$$

This second approximation uses two singular values, 1.30 and 0.83, with two corresponding combinations of stock weights in $\boldsymbol{U}$ and $\boldsymbol{V}$ (compared to the first approximation which has only one singular value). In our new case, $\boldsymbol{U}$ and $\boldsymbol{V}$ become matrices, but the first row of $\boldsymbol{U}$ and first column of $\boldsymbol{V}$ are identical with those in the previous section because they correspond to the same singular value as the first approximation. The second approximation uses $10 + 2 + 12 = 24$ values to represent the 30 entries in $\boldsymbol{W}$.

Now we obtain the approximation:

$$\tilde{W}^{(\mathbf{2})} = \begin{bmatrix} 0.405 & 0.390 & 0.330 & 0.316 & 0.236 & 0.322 \\ 0.558 & 0.561 & 0.043 & 0.014 & 0.204 & 0.334 \\ 0.049 & 0.047 & 0.036 & 0.034 & 0.027 & 0.038 \\ 0.040 & 0.005 & 0.630 & 0.642 & 0.200 & 0.191 \\ 0.049 & 0.047 & 0.036 & 0.034 & 0.027 & 0.038 \end{bmatrix}.$$

Let us again compare the stock weights held by Fund 1 in $\boldsymbol{W}$, which is given by 0.50 on Stock 1 and 0.50 on Stock 2 with the approximation from $\tilde{W}^{(\mathbf{2})}$:

	Stock 1	Stock 2	Stock 3	Stock 4	Stock 5
Fund 1 $\tilde{W}^{(2)}$ weights	0.405	0.558	0.049	0.040	0.049
Actual weights	0.5	0.5	0	0	0

This second approximation much comes closer to the weights in $\boldsymbol{W}$, which has 0.5 weights on both Stocks 1 and 2. The new approximation has only small positions in Stocks 3 to 5.

It is conventional to order the singular values in each approximation from large to small. In this second approximation, the second singular value is 0.83, smaller than the first singular value of 1.30. The more singular values we use, the closer we come to the original $\boldsymbol{W}$. (Each singular value along the diagonal of $\boldsymbol{S}$ is given by the square roots of the eigenvalues.) We can construct a metric measuring commonality, or crowdedness, using the fraction of variance explained by the sum of the first q eigenvalues (here, $q = 2$ for this second approximation), compared to the sum of all eigenvalues required to take us back to the data matrix $\boldsymbol{W}$.

The second combination of funds in the second column of $\boldsymbol{U}$ is given by $\boldsymbol{U}_{:,2} = [-0.090\ -0.650\ -0.016\ 0.754\ -0.016]^T$, which can be interpreted as follows. We observe that the first column of $\boldsymbol{U}$ gives us approximately the rescaled combination of average stock weights held by all funds. Relative to this average weight, the second combination of stocks has relatively large weights in absolute value on Stock 2 (−0.650) and Stock 4 (+0.754). Notice that in the actual $\boldsymbol{W}$ matrix, these are two stocks that the funds tend to hold. Furthermore, funds that tend to hold large amounts of Stock 2, like Funds 1 and 2, tend to hold small amounts of Stock 4. Conversely, funds that tend to have large weights on Stock 4, like Funds 3 and 4, tend to have small weights on Stock 2. Thus, the second column of $\boldsymbol{U}$ has opposite signs on Stocks 2 and 4.

The second column of $\boldsymbol{U}$ is also orthogonal to the first—it is the next linear combination of stocks that is uncorrelated with our first approximation (the first column of $\boldsymbol{U}$), that brings us closer to recovering $\boldsymbol{W}$. (Mathematically, we have $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.) Economically, the second column of $\boldsymbol{U}$ captures a series of salient combination of the stocks relative to the first row of $\boldsymbol{U}$, which has the second-to-best explanatory power, to approximate $\boldsymbol{W}$.

To interpret $\boldsymbol{V}$, we note that each row of $\boldsymbol{V}^T$ represents a combination of "typical" funds. In our first approximation, or in the first row of $\boldsymbol{V}^T$, we effectively uncovered an approximate average weight on funds. The second row of $\boldsymbol{V}^T$ is given by $\boldsymbol{V}^T_{2,:} = [-0.455\ -0.483\ 0.506\ 0.541\ -0.004\ -0.125]^T$. This represents a "barbell" portfolio, which emphasizes the difference between Funds (1 and 2) relative to Funds (3 and 4). In our full $\boldsymbol{W}$, Funds 1 through 4 hold a subset of all stocks (Stocks 1 through 4, with no Stock 5). This combination emphasizes set of funds that exhibit differences among this subset. Relative to the first row of $\boldsymbol{V}^T$, we have identified the next combination of funds that bring us closer to matching $\boldsymbol{W}$; the second row of $\boldsymbol{V}^T$ is the next level of fund differentiation.

3.3 Full decomposition

Finally, we state the full singular value decomposition of $\boldsymbol{W}$:

$$W = \boldsymbol{USV}^T, \tag{2}$$

where

$$\boldsymbol{W} = \begin{bmatrix} 0.5 & 0.2 & 0.4 & 0.2 & 0.2 & 0.5 \\ 0.5 & 0.7 & 0 & 0.1 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0 & 0.2 & 0.1 \\ 0 & 0.1 & 0.6 & 0.7 & 0.2 & 0.1 \\ 0 & 0 & 0 & 0 & 0.2 & 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.633 & -0.090 & 0.718 & -0.276 \\ 0.541 & -0.650 & -0.529 & 0.077 \\ 0.073 & -0.016 & 0.194 & -0.676 \\ 0.544 & 0.754 & -0.362 & 0.063 \\ 0.073 & -0.016 & 0.194 & 0.676 \end{bmatrix}$$

$$\times \begin{bmatrix} 1.30 & 0 & 0 & 0 \\ 0 & 0.83 & 0 & 0 \\ 0 & 0 & 0.42 & 0 \\ 0 & 0 & 0 & 0.27 \end{bmatrix}$$

$$\times \begin{bmatrix} 0.451 & 0.430 & 0.446 & 0.432 & 0.287 & 0.380 \\ -0.450 & -0.483 & 0.506 & 0.541 & -0.004 & -0.125 \\ 0.225 & -0.626 & 0.167 & -0.387 & 0.102 & 0.609 \\ -0.364 & 0.018 & -0.267 & -0.014 & 0.890 & 0.069 \end{bmatrix}$$

3.4 Explaining common components

There are four singular values in $\boldsymbol{S}$ in Equation (2), with each singular value being a square root of the matrix eigenvalue. We can state the proportion of variance explained by the first approximation, which is given by

$$\frac{(1.30)^2}{(1.30)^2 + (0.83)^2 + (0.42)^2 + (0.27)^2} = 65\%.$$

Adding more combinations of stocks, through the columns of $\boldsymbol{U}$, and the combinations of funds, through the columns of $\boldsymbol{V}$, results in closer and closer approximations. The proportion of variance explained by taking two singular values is

$$\frac{(1.30)^2 + (0.83)^2}{(1.30)^2 + (0.83)^2 + (0.42)^2 + (0.27)^2} = 91\%.$$

This ratio of eigenvalues (or the squared terms of the singular values) to total variance is a measure of crowding in stock positions because it captures the most common stock positions across all funds. In the full singular value decomposition, we uncover the full matrix $\boldsymbol{W}$. In actual data, like our numerical example, we can characterize large dimensions of fund types or stock-level positions with only a few singular values.

3.5 Economic interpretation

The singular value decomposition leads to important economic insights. First, it highlights the true dimensionality of the holdings matrix, which has implications pricing. Second, it captures the salient characteristics of stocks (through $\boldsymbol{U}$) or funds (through $\boldsymbol{V}$) in relation to fund data (assets under management, fees, etc.) and stock characteristics (volatility, size, etc.) that can yield insights into crowding.

Beginning with dimensionality, the fact that there are five stocks but only four singular values means that the weight matrix $\boldsymbol{W}$ is not full rank. (The rank of $\boldsymbol{W}$ is four yet there are five stocks.) That is, even though there are more funds than stocks, we do not have a combination of funds that can simultaneously price each individual stock. In our example, Stocks 3 and 5 are not held separately by a single fund, and thus there is no combination of funds for which we can infer a unique price for Stocks 3 or 5. In data, a necessary but not sufficient condition is that we require at least as many funds as stocks for there to be equivalent information in the funds as stocks—however, we realize that only small fluctuations in weights will lead to $\boldsymbol{W}$ to have full rank in a statistical sense. What is more relevant is the effect of the most significant singular values. We estimate singular values with large explanatory power, and conversely that account for only a small fraction of variance, using recent advances in Random Matrix Theory.[9] Changes over time in the fraction of variance explained by the largest singular values provide natural metrics of crowding.

We can also economically interpret the left and right singular matrixes to observed stock and fund characteristics to understand the drivers of decisions by fund managers and investors. For example, suppose the funds have expense ratios (in percent) given by $\boldsymbol{f}^{\mathbf{T}} = [1.1\ 0.9\ 0.5\ 0.5\ 0.1\ 0.15]^T$ and we correlate $\boldsymbol{V}^T$ with $\boldsymbol{f}$ to measure the relation between the canonical loading and expenses across funds. In our numerical example, this yields an R^2 of $0.75^2 = 0.56$. This reflects the fact that the higher cost funds (1 and 2) have greater representation than the lower cost Funds, 5 and 6, because they span the space of stocks. One can similarly—noting that the sign is arbitrary—examine the loadings on the significant eigenvectors across stocks and relate them to characteristics such as firm size or volatility.

In summary, the singular value decomposition approach summarizes data matrixes of fund holdings, expressed in weights or dollars. The procedure uncovers the weights of individual stocks or combinations of funds which characterize the most common components across the fund holdings. The larger the explanatory power of a given set of combinations of stocks or funds, the greater the commonality in positions across the funds, giving us a natural metric for fund crowding over time.

4 Empirical Analysis

4.1 Data and procedures

The data are obtained from Morningstar and consists of quarterly US-listed active equity mutual fund and ETFs holdings from January 1, 2007 to December 31, 2018. We focus on domestic equity funds and exclude leveraged/inverse funds and funds that blend equity and other asset classes. Some funds, despite their domestic equity description, hold foreign stocks, cash, and treasury or corporate bonds and some equity holdings are also not public, such as pre-IPO shares or private equity. To restrict attention to those funds focused on publicly listed US equities, we require that at least 80% of a fund's holdings comprise the Russell 3000 universe. Our results are, as we discuss later, robust to changes in this threshold figure and to the use of other index definitions including the S&P 500. Many mutual funds offer multiple share classes which are simply clones of the same fund, *albeit* with different expense ratios, so we conduct our analysis at the master fund level and using the lowest expense ratio across share classes.

Using these data, we can construct the holdings matrix for active mutual funds and ETFs by quarter for 2007 to 2018. Exhibit 2 lists summary statistics on the number of funds and assets in our universe by year (based on the last quarter's data of each year), which shows the remarkable rise in the number of ETFs over the sample. We note that the number of distinct stocks has been relatively constant in recent years (see Exhibit 1), of which very roughly 3,000 can be considered liquid in terms of spreads and daily volume. Note that our sample is somewhat smaller from the one shown

Exhibit 2 Summary statistics on number of funds and AUM.

	Mutual funds		ETFs	
Year	Number of MFs	AUM ($tn)	Number of ETFs	AUM ($tn)
2007	1,248	1.958	236	0.337
2008	1,305	1.212	254	0.279
2009	1,361	1.592	256	0.337
2010	1,403	1.997	299	0.441
2011	1,504	2.021	374	0.472
2012	1,583	2.349	349	0.582
2013	1,581	3.114	380	0.915
2014	1,699	3.480	396	1.160
2015	1,710	3.364	481	1.195
2016	1,728	3.475	544	1.466
2017	1,667	3.823	626	1.900
2018	1,666	3.406	650	1.849

The exhibit shows the number of funds and assets (in trillions of dollars) by year from January 1, 2007 to December 31, 2018 for the universe of US-listed, US-focused active equity mutual funds and equity ETFs. Funds and AUM are reported at the end of each calendar year.

in Exhibit 1 because of the focus on US equity funds but is still representative of the vast majority of mutual fund and ETF assets domestically.

We analyze the most recent quarter (ending December 31, 2018) in more detail to gain insight. At this time, we have 1,666 US-listed US-focused equity active mutual funds in 2,945 distinct stocks. (There are a few stocks less than the full 3,000 because of corporate actions and mergers.) These active funds collectively have \$3.41 trillion under management. The distribution of fund assets is right skewed; the mean is \$2.04 billion while the median fund has \$369.9 million. The mean expense ratio for the active funds in this quarter is 98 basis points, which is also close to the median, and weighted by AUM, the expense ratio is 0.70%.[10] Consistent with previous research (see, e.g., Carhart, 1997; Fama and French, 2010), the correlation between expenses and AUM is negative at −0.20.

As we would expect, many funds share the same benchmarks: 402 funds are benchmarked to the S&P 500 Index, 307 funds to the Russell 1000 Growth Index, 241 funds to the Russell 1000 Value Index, and 166 funds to the Russell 2000 index. Perhaps surprising, given the large number of securities in these benchmarks, many funds are quite concentrated in terms of numbers of securities held. Indeed, the mean count of holdings (not reported in Exhibit 2) is 132 securities and the median holding is just 67 stocks. Of course, it is possible that these funds may have tight tracking error to their benchmarks if they are sufficiently diversified across sectors/industries.

There are also 650 US-listed, US-focused equity ETFs at December 2018 holding 2,990 distinct Russell 3000 stocks. The total AUM for these ETFs is \$1.85 trillion, with a mean of \$2.84 billion. The top 10 ETFs alone account for \$822.2 billion in assets or 45% of the total. In comparison to the active mutual funds, the mean expense ratio for ETFs is only 41.3 basis points, and it is only 14.2 basis points on an asset-weighted basis. In contrast to the relatively concentrated positions for many mutual funds, most ETFs have diverse holdings: the average ETF holds 274 stocks, and there are 108 funds with 500 plus constituents accounting for over \$1.3 trillion in assets.

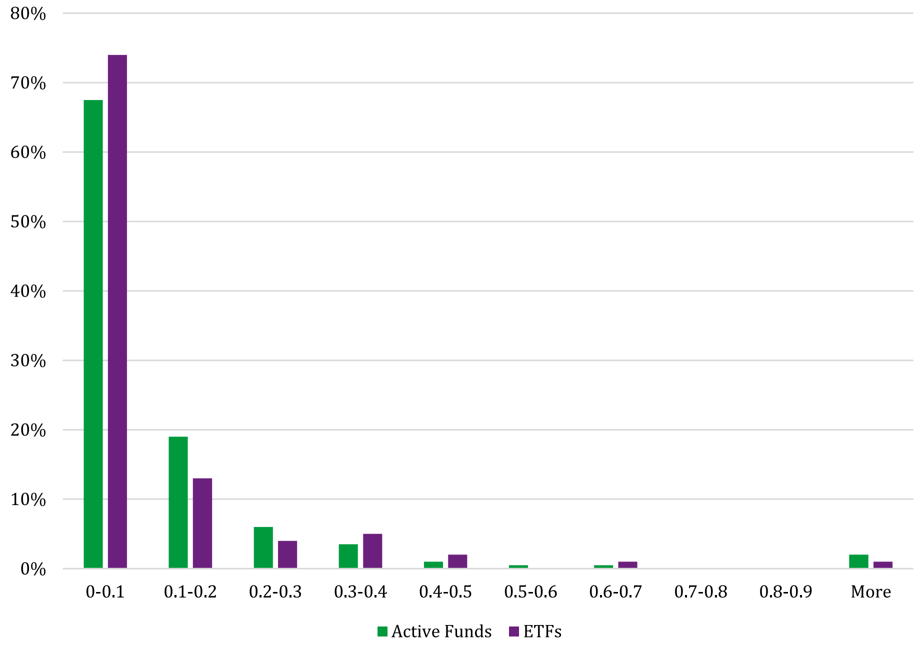
4.2 *Significant singular values and dimensionality reduction*

We start by characterizing the most important singular values by the Marčenko–Pastur distribution of eigenvalues using Random Matrix Theory (Marčenko & Pastur, 1967). This statistical test formally shows that the large dimension $\boldsymbol{W}$ matrix for active funds (with almost 4.9 million entries) can actually be efficiently summarized by a much lower dimension matrix with only 10–15 singular values and their associated eigenvectors. Similar observations hold for the ETF sample (650 funds; 2,990 stocks). This is true not only of the most recent quarter but also of all the quarters in our universe. For active funds as a group, the few largest singular values (the 16 that are above 0.25) are clear statistical outliers. These account for the great majority of the variance, with the smaller singular values (between 0 and 0.25) reflecting noise.

***Observation*:** The holdings of ETFs and active mutual funds across US stocks can be efficiently summarized by approximately 10 canonical funds.

Exhibit 3 shows the frequency distribution of the top 100 singular values for the 1,666 active equity mutual funds and the 650 equity ETFs in our sample on 12/31/2018. Observe that the distributions for both fund vehicles are highly right skewed. Comparing ETFs and active mutual funds, the fraction of singular values below 0.1 for active mutual funds is 68% versus 74%. Correspondingly, the top singular values for active funds are

Exhibit 3 Histogram of largest 100 singular values for active equity mutual funds and ETFs as of 12/31/2018.



larger than the corresponding figures for ETFs as shown below:

Largest singular values	
Active funds	ETFs
5.425	1.677
1.771	0.669
1.019	0.468
0.610	0.402
0.523	0.388

(This is true of other quarters too, as we discuss in the following section.) Thus, we observe:

***Observation*:** There is more commonality explained by the first few canonical funds for active mutual funds versus ETFs.

Having shown that approximately 10 singular values are sufficient to characterize the large fund universe, with there being more commonality in the first few singular values for mutual funds versus ETFs, we have a solid basis for using these statistics to develop crowding measures, over time and by investment styles, measured by holdings corresponding to the top 10 singular values.

4.3 Crowding over time

Exhibit 4 shows the fraction of variance explained by the first 10 singular values for mutual funds, both equally weighted (Panel A) and AUM weighted (Panel B).[11] We observe that commonality among active equity mutual funds has been relatively *constant* over time but increases when weighted by AUM. We perform a similar analysis for ETFs in Exhibit 5, which shows the fraction of variance explained by the first 10 singular values for ETFs, both equally weighted (Panel A) and AUM weighted (Panel B). The percentage of variance explained by the first 10 singular values has been materially *decreasing* in recent years (i.e., from 53–55% in 2007–2008 to 46–48% in 2016–2018). This is consistent with new ETFs

Exhibit 4 Singular variance decomposition for active mutual funds.

	Fraction of variance explained by each singular value (SV)										
Year	SV1	SV2	SV3	SV4	SV5	SV6	SV7	SV8	SV9	SV10	First 10 SVs
Panel A: Equal weight											
2007	0.10	0.05	0.04	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0.31
2008	0.11	0.05	0.04	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.30
2009	0.13	0.05	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.31
2010	0.12	0.05	0.04	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.30
2011	0.13	0.06	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.32
2012	0.13	0.06	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.31
2013	0.11	0.05	0.04	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.29
2014	0.11	0.04	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.26
2015	0.12	0.05	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.28
2016	0.13	0.05	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.29
2017	0.14	0.05	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.29
2018	0.15	0.05	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.31
Panel B: Asset weighted											
2007	0.21	0.06	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.44
2008	0.21	0.07	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.46
2009	0.21	0.08	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.48
2010	0.19	0.08	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.46
2011	0.22	0.10	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.49
2012	0.21	0.09	0.04	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.50
2013	0.21	0.09	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.46
2014	0.18	0.08	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.45
2015	0.21	0.09	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.48
2016	0.23	0.09	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.49
2017	0.26	0.08	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.50
2018	0.30	0.08	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.54

The exhibit shows the fraction of variance in holding of US-listed, US-focused active equity mutual funds explained by the first 10 singular values, by year, based on quarterly holdings data from January 1, 2007 to December 31, 2018. Panel A shows equal weighted results, while Panel B is AUM weighted.

tracking different and niche benchmarks as ETF assets have grown.

***Observation*:** Commonality among equity mutual funds has remained approximately constant, but there has been increased dispersion in ETF offerings.

Comparing Exhibits 4 and 5, we see that the active mutual fund industry is different to ETFs in that the percentage of variance explained by the top 10 eigenvectors, at 37%, for active mutual funds, is much lower compared to ETFs, at 54%. Furthermore, we see no apparent rise in concentration or crowding over the period from January 2007 to December 2018.

***Observation*:** Fewer canonical funds are required to approximate ETFs than mutual funds.

Exhibit 5 Singular value decomposition for ETFs.

	Fraction of variance explained by each singular value (SV)										
Year	SV1	SV2	SV3	SV4	SV5	SV6	SV7	SV8	SV9	SV10	First 10 SVs
Panel A: Equal weight											
2007	0.11	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.45
2008	0.11	0.05	0.05	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.43
2009	0.10	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.43
2010	0.10	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.41
2011	0.12	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.02	0.02	0.43
2012	0.11	0.05	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.40
2013	0.10	0.06	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.41
2014	0.11	0.06	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.41
2015	0.10	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.39
2016	0.10	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.38
2017	0.11	0.05	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.40
2018	0.12	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.39
Panel B: Asset weighted											
2007	0.35	0.17	0.09	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.83
2008	0.40	0.11	0.09	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.84
2009	0.32	0.15	0.08	0.06	0.05	0.04	0.03	0.03	0.03	0.02	0.83
2010	0.30	0.16	0.08	0.06	0.05	0.03	0.03	0.03	0.03	0.02	0.79
2011	0.33	0.16	0.07	0.06	0.05	0.04	0.04	0.03	0.02	0.02	0.82
2012	0.33	0.15	0.08	0.06	0.05	0.04	0.04	0.03	0.02	0.02	0.80
2013	0.31	0.14	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.02	0.76
2014	0.31	0.12	0.07	0.06	0.05	0.05	0.03	0.03	0.03	0.02	0.77
2015	0.32	0.12	0.07	0.06	0.05	0.04	0.04	0.03	0.02	0.02	0.76
2016	0.31	0.12	0.08	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.76
2017	0.35	0.12	0.08	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.77
2018	0.36	0.11	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.02	0.77

The exhibit shows the fraction of variance in holding of US-listed, US-focused ETFs explained by the first 10 singular values, by year, based on quarterly holdings data from January 1, 2007 to December 31, 2018. Panel A shows equal weighted results, while Panel B is AUM weighted.

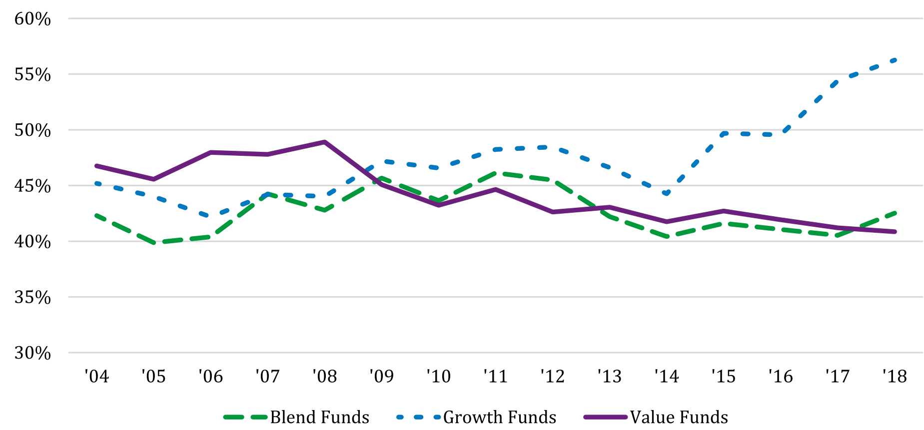
Finally, we also observe that the fraction of variance explained by the first 10 singular values, and the first singular value, is larger for both mutual funds and ETFs moving from equal weights to AUM weights. By construction, this reflects the effect of the largest mutual funds and ETFs. The increase, however, is more pronounced for ETFs because the largest ETFs are tracking broad index benchmarks with low levels of active risk. To summarize:

***Observation*:** The largest ETFs exhibit more commonality than the largest mutual funds because they hold more diversified portfolios.

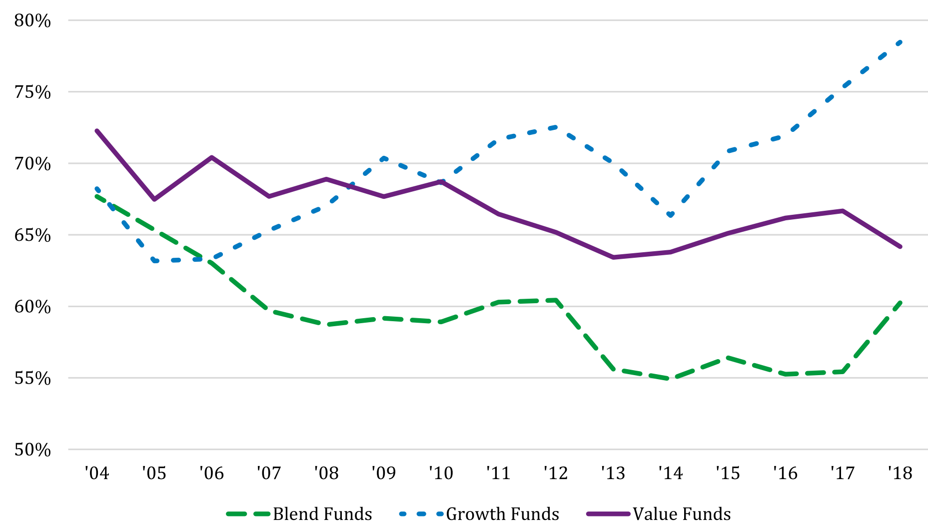
Bhattacharya and O'Hara (2016) argue that rising ETF assets has led to more crowding. On the

Exhibit 6 Crowding in growth, blend, and value funds.

Panel A: Equal weight



Panel B: AUM weight



The charts show the fraction of variance explained by the top 10 singular values by Morningstar fund category for the period 2004–2019. Panel A shows equal weight and Panel B is AUM weighted.

other hand, authors like Glosten *et al.* (2017) find the creation of ETFs positively increases information efficiency at the stock level. The statistics we cite here describe commonality among ETFs and mutual funds, rather than the largest absolute positions. The highest percentage ownership by ETFs is in US small-capitalization stocks, North American gold miner stocks, and the Nikkei 225 Index, the latter reflecting Bank of Japan holdings. Overall, index ownership remains small.

4.4 Crowding across style factors

We examine commonality in active mutual funds in Morningstar style box categories in Exhibit 6. We see that while Value and Blend commonality has remained roughly constant over the period 2007–2018, there has been a marked increase in our crowding metric for growth funds:

***Observation*:** Growth funds exhibit the largest increase in crowded positions.

Exhibit 7 Singular value decomposition for ETFs in factor space.

Year	Number of ETFs	ETF AUM ($Tn)	Fraction of variance (in %) explained by each principal component (PC)											
			Rank of X	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	First 10 PCs
Panel A: ETF factor SVD (BlackRock Fundamental Equity Risk Model—US/Americas Region)														
2017	663	1.760	60	52.6	20.0	9.2	4.9	3.6	2.0	1.5	1.2	1.0	0.8	96.8
2016	589	1.336	60	54.6	15.3	8.8	7.4	2.7	2.2	1.9	1.4	1.2	0.9	96.4
2015	510	1.098	60	52.6	20.9	11.7	3.3	2.3	1.6	1.4	1.1	0.8	0.7	96.3
2014	434	1.072	60	55.1	18.4	11.3	3.6	2.2	1.8	1.5	1.0	0.8	0.7	96.3
2013	405	0.800	60	55.9	18.7	9.8	3.3	2.7	2.5	1.2	1.0	0.8	0.5	96.4
Panel B: Dollar (AUM) weighted ETF factor SVD (BFRE USAM Model)														
2017	663	1.760	60	72.4	14.9	5.0	2.1	1.7	1.0	0.7	0.5	0.4	0.3	98.9
2016	589	1.336	60	74.6	11.3	5.6	3.3	1.1	1.0	0.7	0.7	0.5	0.3	99.0
2015	510	1.098	60	72.3	14.7	6.6	1.7	1.0	0.8	0.6	0.5	0.4	0.3	98.9
2014	434	1.072	60	73.3	13.5	6.1	2.1	1.1	1.0	0.7	0.5	0.3	0.3	99.0
2013	405	0.800	60	72.5	14.3	5.3	2.2	1.7	1.4	0.5	0.4	0.3	0.2	98.9

When using AUM weights in Panel B of Exhibit 6, we can see the increased commonality of growth funds even more clearly, with the Blend Funds exhibiting a decrease in commonality.

In Exhibit 7, we report crowding metrics in terms of style factors, as described in the Appendix. Essentially, we permute the weight matrix $\boldsymbol{W}$ in stock space into a holdings matrix in style factor space, where each stock is mapped to a z-score for k factors. We specify factor z-scores based on the universe of all US stocks as regressors: *Size* increases with market capitalization, *Volatility* is the standard deviation of daily returns, Momentum is the 12-month return less the most recent month's return, *Value* is measured by earnings/price and price/book, *Earnings* is earnings yield based on past three-year earnings, and *Dividend* is the three-year dividend yield.

***Observation*:** Crowding is greater in factors than in individual stocks for both active mutual funds and ETFs. The level of factor crowding has remained relatively constant over the period.

4.5 *Characterizing canonical funds*

In the singular value decomposition, the columns of the left singular matrix $\boldsymbol{U}$ are portfolios of stocks, ordered by explanatory power, while the rows of the right singular matrix $\boldsymbol{V}$ are groups of funds that represent "archetypal" portfolios, again ordered by their ability to explain the dispersion in the weight matrix. In this section, we characterize the two most important linear combinations of funds that represent the most common typical fund types.

The interpretation of a large vector is difficult and subjective, but we gain insights by relating the loadings across funds to characteristics of these funds, as we did in the numerical example (see Section 2). We turn first to active mutual funds.

In Exhibit 8, we report cross-sectional regressions of the mutual fund vectors (right singular matrix) corresponding to the largest two singular values on mutual fund-level explanatory variables. These include the expense ratio, log of fund

Exhibit 8 Characterizing canonical mutual funds.

	First eigenvector of funds		Second eigenvector of funds	
	Estimate	t-Value	Estimate	t-Value
(Intercept)	−0.027	−2.78	−0.019	−1.36
Expense ratio (Net)	0.003	3.71	0.002	1.99
log AUM	0.000	−1.78	0.000	2.47
No. of holdings	0.000	−1.18	0.000	−1.88
Russell 1000 Growth	−0.015	−1.60	0.034	2.50
Russell 1000 Value	0.010	1.01	−0.030	−2.24
Russell 2000	0.026	2.65	0.014	1.03
Russell 2000 Growth	0.026	2.63	0.015	1.06
Russell 2000 Value	0.026	2.66	0.014	0.99
Russell MidCap	0.022	2.24	0.010	0.70
Russell MidCap Growth	0.021	2.20	0.016	1.17
Russell MidCap Value	0.023	2.37	0.006	0.43
S&P 500	0.003	0.29	0.000	0.01
Adjusted R^2	0.706		0.668	
F-statistic	334.1		279.8	
Degrees of freedom	1653		1653	

We report regressions of the fund eigenvectors (right singular matrix) corresponding to the largest two singular values for 12/31/2018. There are1,666 active equity mutual funds and 2,946 stocks in the sample weight matrix. Note that the coefficient signs in each regression are arbitrary.

assets, position count, and benchmark, which have been previously used by an extensive mutual fund literature (see, e.g., Carhart, 1997). The coefficient signs within each regression can be reversed because the sign on the vectors in the singular matrices can be flipped so accordingly, we focus on the overall goodness of fit, the statistical significance of the explanatory variables, and any sign differences within the regression.

The regression of the first right singular vector of the right singular matrix $\boldsymbol{V}$ (that corresponds to the largest singular value) offers some interesting insights. Although the explanatory variables were not used in the singular value decomposition, they collectively explain a high fraction of the loadings across funds with an R^2 of 0.71. Clearly, expense ratio matters, as does fund size. Perhaps surprisingly, the right singular vectors do not load significantly on S&P 500 funds possibly because large, blend styles are subsumed in the other benchmark variables—which are statistically significant. (Note that the Russell 1000 is implicit in the intercept term in the regression). Thus, we can interpret the first right singular vector as a weighting on funds that closely track the market.

***Observation*:** The first and most important archetypal "fund" that approximates the whole active fund universe is a market capitalization index.

Note that this result is not by construction for three reasons. First, because funds constitute only a minority of holdings of the complete market (see, e.g., Ang *et al.*, 2017b). Second, funds have tracking error relative to those many different market benchmarks, and finally, the "market" benchmarks to which funds are compared are non-identical and in some cases have significant tracking error relative to each other (like the S&P 500 and Russell 3000).

In the second regression of the second right singular vector, we observe a lower R^2 of 0.67, but also significant and opposing loadings on growth versus value indexes. In other words, the second eigenvector corresponds to a value–growth axis. The value factor is the first factor identified in practitioner treatises (Graham and Dodd, 1934) with the first serious academic work in Basu (1977), and the fact it shows up so clearly in this decomposition shows the pervasive differentiation of mutual funds into growth versus value strategies.

***Observation*:** After market exposure, the second most important form of differentiation for active funds is in value–growth factor exposure.

4.6 *Interpretation of ETFs*

For ETFs as a group, the interpretation of the canonical funds is much easier as we can examine the funds with the highest correlation with the first few singular vectors. The first stock singular vector exhibits a 99% correlation with the S&P 500, and the funds with the highest correlation to this eigenvector are "broad-market" US equity funds. In other words, the first singular vector is the market itself. This is consistent with the history of the industry: the first US-listed ETF was launched in 1993 and the fact that it is only relatively recently that ETFs have seen more niche products. Today, market capitalization ETFs continue to predominate in terms of assets. The second singular vector is more complex and reflects loadings on growth stocks and certain sectors, particularly energy.

***Observation*:** Similar to active funds, the market factor is most important in approximating the space of ETFs. Unlike active funds, the next degree of differentiation reflects both value–growth style factor exposure and sector exposures.

4.7 *Factor characteristics of canonical stock portfolios*

Exhibit 9 reports regressions of the left singular vectors (corresponding to the left singular matrix of stocks, $\boldsymbol{V}$). We take the left singular vectors corresponding to the largest two singular values at December 2018. We specify factor z-scores based on the universe of all US stocks as regressors: *Size* increases with market capitalization, *Volatility* is the standard deviation of daily returns, Momentum is the 12-month return less the most recent month's return, *Value* is measured by earnings/price and price/book, *Earnings* is earnings yield based on past three-year earnings, and *Dividend* is the three-year dividend yield.

In Exhibit 9, we observe significant loadings on *Size* meaning that the first left singular vector is associated with larger stocks. The opposite sign is present on *Value*. Because the singular vectors have arbitrary sign, we can only interpret *relative* differences in the coefficients. A natural interpretation is that the first approximation of the stock portfolios tilt toward larger, growth-oriented stocks. The second left singular vector loads on *Volatility* and *Size*, again with opposite signs, possibly picking up a small cap and high volatility tilt.

***Observation*:** Overall, the first and most important characterization of the stocks held by funds is that the stocks tend to be large, growth stocks. On top of this, funds tend to hold stocks that are smaller with higher volatilities.

4.8 *Market efficiency with more funds than stocks*

Finally, we turn to the question of market efficiency of funds versus stocks, especially in today's world of more funds than stocks. The analysis in this final section is particularly important, given the earlier cited concerns regarding the possible mispricing of stocks due to flows into ETFs, and the possibility that liquidity moves away from

Exhibit 9 Factor regressions of canonical stock portfolios.

	First eigenvector of stocks (×100)		Second eigenvector of stocks (×100)	
	Estimate	t-Value	Estimate	t-Value
(Intercept)	1.660	33.088	1.092	19.861
Size	0.801	29.346	0.482	16.119
Volatility	0.050	1.447	−0.092	−2.422
Momentum	0.007	0.264	−0.105	−3.244
Value	−0.209	−6.469	0.005	0.129
Earning yield	−0.038	−1.244	−0.047	−1.384
Dividend yield	−0.048	−1.379	0.184	4.831
Adjusted R^2	0.251	0.123		
F-statistic	165.6	68.7		
Degrees of freedom	2938	2938		

We report regressions of the stock eigenvectors (left singular matrix) corresponding to the largest two singular values for 12/31/2018. There are 1,666 mutual funds and 2,946 stocks in the sample weight matrix. Note that the coefficient signs in each regression are arbitrary. Eigenvectors are scaled by 100 and are arbitrary in sign. Factors are z-scores based on the universe of all US stocks: *Size* increases with market capitalization, *Volatility* is the standard deviation of daily returns, Momentum is the 12-month return less the most recent month's return, *Value* is measured by earnings/price and price/book, *Earnings* is earnings yield based on the three-year earnings, and *Dividend* is the three-year dividend yield.

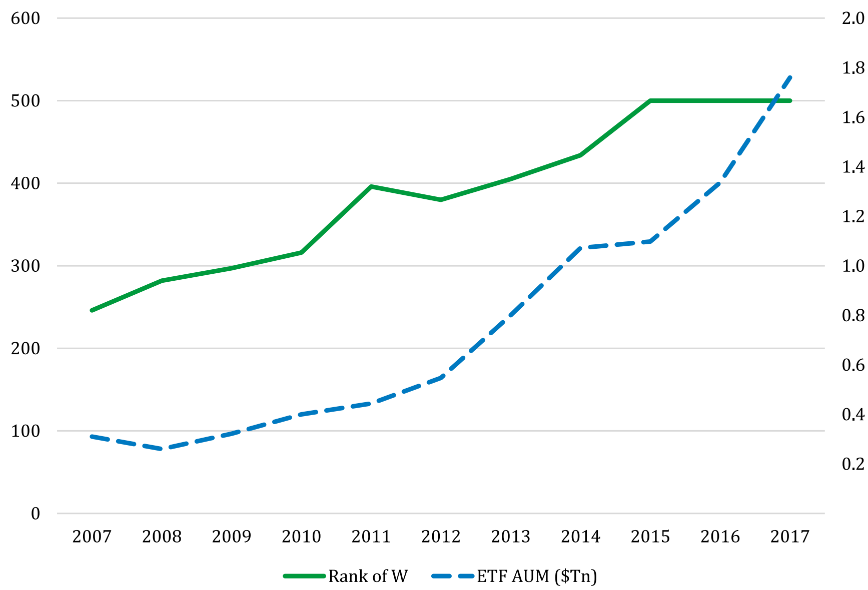
individual stocks to funds (see Brogaard *et al.*, 2020). Similarly, Sağlam *et al.* (2019) argue that stocks with high ETF ownership may experience impaired liquidity during stress events. We show that these concerns are misplaced. Although it is common to view ETFs as index products, they can also be tools for price discovery that are increasingly used by hedge funds and active managers to express views on sectors, industries, countries, regions, capitalization ranges, and factors.[12]

ETF prices are determined on organized exchanges through the interaction of supply and demand. We show that if these prices are correct, then the underlying constituents can be priced from the funds' market determined prices even if there is no trading or liquidity in the underlying stocks. In other words, the fund prices can suffice for price discovery!

We focus on the constituents of the S&P 500 Index, the most important benchmark for large-cap US equities which represents about 80% of the total US stock market.[13] As of May 31, 2019, the S&P 500 total market capitalization is about $24 trillion; approximately $9.9 trillion was benchmarked to the index, with index assets comprising approximately $3.4 trillion.[14] There are many more US-listed, US-focused equity ETFs than the 500 stocks in the index: 1,481 funds in 2018. In our analysis, we do not impose any restriction on "style" (we include all style box, sector, and other style funds) or fund domicile, but we only take funds that are long only. This leaves a sample of 2,844 funds with 500 distinct stocks. We further refine this sample by taking funds that have a majority of their overall holdings in the S&P 500 Index (like US large-cap blend) and create a panel of funds and years, from 2007 to 2018.

In Exhibit 10, we plot the rank of the weight matriculates, $\boldsymbol{W}$. If $\boldsymbol{W}$ is not full rank, then the fund universe is not sufficient to ensure that the underlying stocks are efficiently priced (as in the example in Section 2). If $\boldsymbol{W}$ is full rank, then

Exhibit 10 ETF holding matrix rank.



We report the rank of the matrix of ETFs (left-hand side) and the assets under management (in trillions of dollars) for the S&P 500 universe.

investors are indifferent between price discovery occurring in ETF markets or underlying physical markets—both lead to efficient pricing in stocks. Exhibit 9 shows that as of December 31, 2018, the correlation matrix $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{W}^T$ has had full rank (= 500) and is invertible. But this is only the case since 2015. Before then, there were fewer listed ETFs that held S&P 500 stocks compared to number of stocks in S&P 500 index itself (= 500), and the rank of C was less than 500.

Observation**:** Efficient price discovery of underlying stocks in the S&P 500 can occur through ETFs or through trades of the underlying stocks themselves—but this has only been the case fairly recently since 2015.

5 Conclusion

Today, there are more active mutual funds and ETFs than individual stocks. Given the increasing numbers of funds focused on a stock universe that has decreased in recent years, it could be that the risk of crowding, and associated short-term reversals are perhaps more likely. For active funds, it is possible that certain managers focused explicitly or implicitly on a subset of stocks, represented by certain factors, sectors, or other themes, and other managers might gravitate to the same sorts of positions. For ETFs, the vast majority of which track a given index, the possibility of crowding may arise because the assets under management are determined by investor flows into open-end funds. ETFs can have factor exposures even if they are not factor funds, and even if most of them follow indexes. For example, investors chasing stock returns may lead to increased crowding in funds exposed to momentum either explicitly (through factor momentum funds) or implicitly through funds with similar positions that have seen high returns.

Using singular value decompositions with time-series of all individual stock weights of all funds from 2007 to 2018—for both mutual funds and ETFs—we characterize commonality across stocks held by funds. We also characterize commonality across fund types and construct canonical portfolios that capture the commonality across positions. An important result is that efficient pricing can be obtained by trading funds even if the underlying stocks were not to trade—there are enough stocks held by these funds, in the right combinations, that results in efficient prices for the underlying stocks. This is only true since 2015. Since the numbers of mutual funds have remained steady since this time, this fact is contemporaneous with the rapid entry of new ETFs.

We show that despite the large number of funds, fewer than a dozen fund canonical portfolios ("archetypes") well describe the whole universe of funds. The commonality of active mutual funds has remained approximately constant in the period 2007 to 2018, with decreases in common ETF positions over this time. Mutual fund positions are less alike than ETFs; fewer canonical fund types are required for good approximations for ETFs than for active mutual funds. This is consistent with the proliferation of ETFs tracking various indexes. Looking at style box categorizations, we find evidence of increased crowding over the last decade among growth-oriented mutual funds although that factor crowding in common factor exposures has remained relatively constant. In conclusion, fears that the proliferation of active and index funds has increased crowding are not supported by the empirical evidence.

Appendix

Consider a weight matrix, $\boldsymbol{W}$, with dimension $n \times p$, where n is the number of stocks in the universe and p is the number of funds (portfolios). Throughout we will use uppercase letters to denote matrixes and lower case letters to denote scalars. Each column of $\boldsymbol{W}$ represents a particular fund's portfolio weight (or assets in dollars) across the universe of stocks. Specifically, element (i, j) of $\boldsymbol{W}$ is the weight (a number between 0 and 1) on stock $i = 1, \ldots n$ by fund $j = 1 \ldots p$. We will compute $\boldsymbol{W}$ separately for active funds and ETFs, as well as jointly.

Since many funds may have broad market indexes such as the S&P 500 as their benchmark, we might expect these portfolio weights to be very similar. Similarly, a Russell 3000 fund's positions are the sum of a Russell 1000 and Russell 2000 funds, by construction. A logical question is: "What is the true dimension of the holdings matrix?" In a mathematical sense, we can ask if some of the columns of $\boldsymbol{W}$ are linear combinations of other columns, including the case where two or more funds hold identical portfolios. That is determined by the rank of the matrix, the largest linearly independent subset of columns.

For active equity funds, matrix $\boldsymbol{W}$ has full rank, meaning that no two funds in our sample have the exact same holdings, which is consistent with security selection by actively managed funds. Interestingly, this is only true relatively recently, as Section 4.6 discusses.

Another method to characterize $\boldsymbol{W}$ is to reduce dimensionality and approximate $\boldsymbol{W}$ by only a few portfolios held by a few canonical fund types. Formally, we seek to summarize the variation in holdings across the p portfolios with a much smaller group of funds $q < p$ for any q.

Singular value decomposition is a useful technique to reduce dimensionality. The Singular Value Decomposition of $\boldsymbol{W}$ factorizes that matrix into the product of three matrices:

$$W = \boldsymbol{USV}^T. \tag{A.1}$$

Here,

- $\boldsymbol{U}$ is an orthogonal matrix[15] with dimension $(n \times n)$, termed the *left singular matrix*;

- $\boldsymbol{S}$ is a matrix with dimension $n \times p$ whose non-diagonal entries are zero and whose diagonal entries are non-negative; and
- $\boldsymbol{V}$ is an $p \times p$ orthogonal matrix, and $\boldsymbol{V}^T$ denotes its transpose, termed the *right singular matrix*.

The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal (that is $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix) and the matrix $\boldsymbol{S}$ has positive real entries along the diagonal, arranged from largest to smallest known as singular values, with zero entries elsewhere. The first singular value σ_1 is the square root of the first eigenvalue $\sigma_1 = \sqrt{\lambda_1}$, followed along the diagonal of $\boldsymbol{S}$ by the square roots of the other eigenvalues. In economic terms, the columns of $\boldsymbol{U}$ are long–short portfolios of stocks, ordered by explanatory power, while the rows of $\boldsymbol{V}$ are groups of funds that represent "archetypal" portfolios, again ordered by their ability to explain the dispersion in the weight matrix. Finally, the matrix $\boldsymbol{S}$ is useful in understanding how far we can go in terms of reducing the dimensionality of the weight matrix with a limited amount of information.

Dimensionality reduction

From the singular value decomposition of $\boldsymbol{W}$, we can get a unique matrix of rank $q < p$ that best approximates $\boldsymbol{W}$ for every q. Specifically, in Equation (A.1) we keep only the top q singular values in the matrix $\boldsymbol{S}$ to get a $q \times q$ diagonal matrix $\mathbf{S}_q$ whose non-diagonal entries are zero. Similarly, we keep only the top q left singular vectors (first q columns of $\boldsymbol{U}$) and create an $n \times q$ matrix $\boldsymbol{U}_q$ and likewise keep the top q right singular vectors (first q rows of $\boldsymbol{V}^T$) and create a $q \times p$ matrix $\boldsymbol{V}_q^T$. Then, $\boldsymbol{W}_q = \boldsymbol{U}_q\boldsymbol{S}_q\boldsymbol{V}_q^T$ is a reduced rank representation of the original weight matrix $\boldsymbol{W}$. Formally, it can be proved that $\boldsymbol{W}_q$ is the closest in matrix "distance" to $\boldsymbol{W}$ over all possible rank q matrices $\boldsymbol{A}$ with dimensions $n \times p$:

$$\|\boldsymbol{W} - \boldsymbol{W}_q\|_F \leq \|\boldsymbol{W} - \boldsymbol{A}\|_F \qquad \text{(A.2)}$$

where $\|\cdot\|_F$ denotes the distance between two matrices.[16] SVD is not a parametric approach like a linear regression; rather, SVD finds another basis, which is a linear combination of the original basis that best expresses the dispersion in the original data.[17] From an economic perspective, this means that we can quantify how many distinct funds are needed to approximate of the variation in the weight matrix as closely as we would like. The original weight matrix has np entries, while the reduced rank representation using SVD requires $q(n + p + 1)$ numbers. When q is much less than n, this can be a considerable reduction in dimensionality.

Our crowding metric at any point in time is the fraction of variance explained by the first q-squared singular values:

$$\Gamma_t^q = \left(\frac{\lambda_1 + \cdots + \lambda_q}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}\right) \qquad \text{(A.3)}$$

We provide statistics for this metric of similarity in fund holdings for all mutual funds and ETFs, separately, and for individual stocks and for factors.

Factor representation

To understand the commonality in factor space, we assume a standard multi-factor model where the return of stock s at time t in excess of the risk-free rate, $r_{s,t}$, is a linear function of k factor returns—denoted by $F_{j,t}$ for $j = 1, \ldots, k$—with betas that vary over time:

$$r_{s,t} = \alpha_s + \sum_{j=1}^{k} \beta_{s,j,t-1} F_{j,t} + \varepsilon_{s,t}, \qquad \text{(A.4)}$$

where $\beta_{s,j,t-1}$ denotes the exposure of stock s to factor j at time t. We instrument the beta $\beta_{s,j,t-1}$ for returns at time t to emphasize that it is measurable with respect to information at time $t-1$. We summarize this information in a characteristics vector, denoted by $z_{s,t-1}$. The constant or alpha is not time subscripted, meaning it does not itself vary with time, but returns in excess of the time-varying factor exposures are subject to stochastic shocks, $\varepsilon_{s,t}$.

We use holdings data, following authors like Gagliardini *et al.* (2016) to estimate factor exposures. This uses more information than traditional time-series regressions (see also Ang *et al.*, 2017a) and can capture time-varying factor loadings; time-series regressions, by assumption, assume the factor exposures are constant over the period used to run the regressions.

Let $\boldsymbol{B}(z_{t-1})$ denote the factor exposure matrix across stocks and factors, given the stock-specific characteristics we observe. This matrix has dimension $n \times k$, and each row represents the exposure of a stock to the k factors in Equation (A.4). If, say, the first of the k factors is "value" as measured by book-to-market and other such metrics, the first column of $\boldsymbol{B}(z_{t-1})$ corresponds to the cross-sectional z-score with respect to value. We can transform the portfolio weight matrix into a factor weight matrix $\boldsymbol{X}$ of dimension $k \times p$ defined as:

$$\boldsymbol{X} = \boldsymbol{B}(z_{t-1})^T \boldsymbol{W} \tag{A.5}$$

Each column of $\boldsymbol{X}$ represents a fund's exposure to each of the k factors. We then reframe the dimensionality reduction question posed earlier to fund factor exposures using the singular value decomposition on $\boldsymbol{X}$. This allows us to understand crowding in factor space, for example, with regard to popular strategies.

Notes

1 See Gao *et al.* (2013) and Doidge *et al.* (2017), among others. The number of US exchange-listed companies at year-end 2018 was 4,397, with the peak of stock listings was in 1996 when 8,090 stocks. The number of equity funds first exceeded the number of stocks in 2007. As of year-end 2018, there were 1,510 ETFs and 4,753 mutual funds listed in the US that were purely equity-focused. Restricting attention to purely US-focused, US-listed equity funds, we find rough parity with the number of US-listed stocks as of 2018, with 4,068 stocks and 4,397 funds. (*Source*: ICI and World Bank, as of December 2018.)

2 Indeed, since the aggregation of index and active equity funds is the market as a whole, and index funds largely seek to track the market, it must imply that the set of active funds must look much like the market.

3 This approach is common in textual analysis and in medical research. In analyses of texts, a researcher may create a word and documents matrix, where each row is a word and each column the word count in a given document. The information in this matrix is summarized by broad themes or stories. In a medical example, the rows may represent genes, and the columns individuals or trials, with the goal of identifying clusters genes that may play a role in causing cancer.

4 Data on the number of US stocks is drawn from the World Bank (2019); data on funds is sourced from the ICI Fact Book (2019). The numbers for ETFs and the mutual fund totals include only pure US-focused equity funds domiciled in the US, and exclude international/global equity funds, commodity funds, bond funds, money market funds, and hybrid funds.

5 A large literature documents the increasing popularity of ETFs, along with concerns about the impact of flows into indexing more broadly. See Madhavan (2016) for a summary of these issues.

6 Source: World Federation of Exchange Database, HFR, Cerulli, Simfund, and McKinsey Cube data (as of December 31, 2018). Equity and Fixed Income exposures estimates are based on individual holdings-level data for US mutual funds and exchange-traded portfolios (which include ETFs).

7 In fact, Berk and Green (2004) present a rational model where fund managers add no economic value beyond the returns of the underlying stocks, but there are endogenous flows to different funds. For other references, see Brown *et al.* (1996), Lynch and Musto (2003), and Evans (2010).

8 Of course, the row sums across columns need not add to 1.

9 Marčenko and Pastur (1967) predict the theoretical upper and lower bounds on the null distribution of eigenvalues for a random matrix. Significant principal

components are those with eigenvalues greater than the maximum eigenvalue predicted for random data, where the data has been centered and scaled. The Tracy–Widom distribution provides the distribution of the largest eigenvalues. A survey of the advances in the area is in Anderson *et al.* (2010).

10 Throughout this paper, we use the term "expense ratio" to refer to the net expense ratio.

11 Other authors have used other approaches to measure crowding, like Bruno *et al.* (2018) who emphasize common risk factor signals (commonality in risk models), or Khandani and Lo (2011) in terms alpha or return-enhancing strategies sharing the same signals (commonality in return signals).

12 There are other notions of price efficiency other than the spanning interpretation we use here. For example, efficiency can be also be defined relative to information flows (see Easley *et al.*, 2019) or arbitrage opportunities (see Garleanu and Pedersen, 2011).

13 Due to index reconstitution, additions, spin offs, and other corporate actions, the official number of S&P 500 constituents can actually be slightly larger than 500 in any given month.

14 Source: S&P Dow Jones Indexes, as of May 31, 2019. The amount indexed to the S&P 500 is an estimate based on survey methods by the index provider.

15 A matrix is orthogonal if its columns (or rows) are orthonormal vectors, meaning they all have length 1 and the inner product of any distinct vector pair is 0. Note that orthonormality of the left and right singular vectors mean they do not need to sum to zero.

16 Technically, the *Frobenius norm* of a matrix is defined as the square root of the sum of the squares of the elements. Another common distance metric is the 2-norm which yields the largest singular value. For any matrix A, the sum of squares of the singular values equals the Frobenius norm.

17 Principal components analysis (PCA) and SVD are closely related. PCA works by finding a set of vectors (basis) that are orthogonal "right angle" projections to each other. The principal components are linear combinations of the rows of $\boldsymbol{W}$ that maximize the variance of the holding matrix in each dimension. SVD yields more information than PCA, namely the matrix $\boldsymbol{U}$, helping us understand the column structure of $\boldsymbol{W}$ too. The row structure helps distinguish canonical fund types (e.g., "value"), while the column structure shows what stocks are liked by the various fund archetypes. This additional information is most valuable when we transform raw weights into factor loadings. PCA also operates on the correlation or covariance matrix which, given our comments earlier on sparseness, may introduce computational challenges. By contrast, SVD works directly on the weight matrix.

References

Anderson, G., Guionnet, A., and Zeitouni O. (2010). *An Introduction to Random Matrices*. Cambridge: Cambridge University Press.

Ang, A., Madhavan, A., and Sobczyk, A. (2017a). "Estimating Time-Varying Factor Exposures," *Financial Analyst Journal* **43**, 41–54.

Ang, A., Madhavan, A., and Sobczyk, A. (2017b). "Crowding, Capacity, and Valuation of Minimum Volatility Strategies," *Journal of Index Investing* **7**, 41–50.

Baltas, N. (2019). "The Impact of Crowding in Alternative Risk Premia Investing," *Financial Analyst Journal* **75**, 89–104.

Basu, S. (1977). "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis," *Journal of Finance* **32**, 663–682.

Berk, J. and Green R. (2004). "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy* **112**, 1269–1295.

Bhattacharya, A. and O'Hara, M. (2016). "Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets," https://ssrn.com/abstract=2740699.

Bonne, G., Roisenberg, L., Kouzmenko, R., and Zangari, P. (2018). "MSCI Integrated Factor Crowding Model," Working Paper, MSCI.

Brogaard, J., Heath, D., and Huang, D. (2020). "ETF Trading and the Bifurcation of Liquidity," Working Paper, University of Utah.

Brown, G., Howard, P., and Lundblad, C. (2019). "Crowded Trades and Tail Risk," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3326802.

Brown, K. C., Harlow, W. V., and Starks, L. T. (1996). "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance* **51**, 85–110.

Brown, K. C., Harlow, W. V., and Zhang, H. (2009). "Staying the Course: The Role of Investment Style Consistency in the Performance of Mutual Funds".

Bruno, S., Chincarini, L., and Ohara, F. (2018). "Portfolio Construction and Crowding," *Journal of Empirical Finance* **47**, 190–206.

Carhart, M. (1997). "On Persistence of Mutual Fund Performance," *Journal of Finance* **52**, 57–82.

Connor, G. and Korajczyk, R. (1988). "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics* **21**, 255–289.

Coval, J. and Stafford, E. (2007). "Asset Fire Sales (and Purchases) in Equity Markets," *Journal of Financial Economics* **86**, 479–512.

Daniel, K. and Titman, S. (1997). "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance* **52**, 1–33.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks," *Journal of Finance* **52**, 1035–1058.

Doidge, C., Karolyi, G. A., and Stulz, R. (2017). "The U. S. Listing Gap," *Journal of Financial Economics* **123**, 464–487.

Easley, D., Michayluk, D., O'Hara, M., Patel, V., and Putnins, T. (2019). Information Flows and Asset Pricing. Working Paper, Cornell University.

Evans, R. (2010). Mutual Fund Incubation. *Journal of Finance* **65**, 1581–1611.

Fama, E. and French, K. (1993). "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* **33**, 3–56.

Fama, E. and French, K. (2010). "Luck versus Skill in the Cross-section of Mutual Fund Returns," *Journal of Finance* **65**, 1915–1947.

Gagliardini, P., Ossola, E., and Scaillet, O. (2016). "Time-Varying Risk Premium in Large Cross-Sectional Equity Datasets," *Econometrica* **84**, 985–1046.

Gao, X., Ritter, J. R., and Zhu, Z. (2013). Where Have All the IPOs Gone? *Journal of Financial and Quantitative Analysis* **48**, 1663–1692.

Garleanu, N. and Pedersen, L. H. (2011). "Margin-Based Asset Pricing and Deviations from the Law of One Price," *Review of Financial Studies* **24**, 1980–2022.

Glosten, L., Nallareddy, S., and Zou, Y. (2017). ETF Activity and Informational Efficiency of Underlying Securities. https://ssrn.com/abstract =2846157.

Graham, B. and Dodd, D. (1934). *Security Analysis*. New York, NY: McGraw-Hill.

Hasanhodzic, J. and Lo, A. W. (2007). "Can Hedge-Fund Returns be Replicated?: The Linear Case," *Journal of Investment Management* **5**, 5–45.

Jiang, G., Yao, T., and Yu, T. (2007). "Do Mutual Funds Time the Market? Evidence from Portfolio Holdings," *Journal of Financial Economics* **86**, 724–758.

Khandani, A. and Lo, A. W. (2011). "What Happened to the Quants in August 2007? Evidence from Factors and Transactions Data," *Journal of Financial Markets* **14**, 1–46.

Lynch, A. and Musto, D. (2003). "How Investors Interpret Past Fund Returns," *Journal of Finance* **58**, 2033–2058.

Madhavan, A. (2016). Exchange-Traded Funds and the New Dynamics of Investing, Oxford University press.

Marčenko, V. and Pastur, L. (1967). "Distribution of Eigenvalues for Some Sets of Random Matrices," *Sbornik: Mathematics* **1**, 457–483.

Marks, J. M. and Shang, C. (2019). "Factor Crowding and Liquidity Exhaustion," *Journal of Financial Research* **42**, 147–180.

Pastor, L., Stambaugh, R., and Taylor, L. (2015). "Scale and Skill in Active Management," *Journal of Financial Economics* **116**, 23–45.

Sağlam, M., Tuzun, T., and Wermers, R. (2019). Do ETFs Increase Liquidity? Working Paper, University of Maryland.

Sharpe, W. F. (1992). "Asset Allocation: Management Style and Performance Measurement," *Journal of Portfolio Management* **18**, 7–19.

Stambaugh, R. (2019). "Skill and Fees in Active Management," Working Paper, University of Pennsylvania.

Wermers, R. (2000). "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *Journal of Finance* **55**, 1655–1695.

Wermers, R. (2012). "A Matter of Style: The Causes and Consequences of Style Drift in Institutional Portfolios," papers.ssrn.com/abstract_id=2024259.

Keywords: ETFs; active mutual funds; crowding; factors