PRODUCT DIFFERENTIATION, SEARCH COSTS, AND COMPETITION IN THE MUTUAL FUND INDUSTRY: A CASE STUDY OF S&P 500 INDEX FUNDS*

Ali Hortaçsu and Chad Syverson

October 2003

We investigate the role that non-portfolio fund differentiation and information/search frictions play in creating two salient features of the mutual fund industry: the large number of funds and the sizeable dispersion in fund fees. In a case study, we find that despite the financial homogeneity of S&P 500 index funds, this sector exhibits the fund proliferation and fee dispersion observed in the broader industry. We show how extra-portfolio mechanisms explain these features. These mechanisms also suggest an explanation for the puzzling late-1990s shift in sector assets to more expensive (and often newly entered) funds: an influx of high-information-cost novice investors.

---

* Zvi Eckstein and Alan Sorensen provided thoughtful suggestions regarding earlier drafts. We also thank Judith Chevalier, Lars Hansen, Thomas Hubbard, Boyan Jovanovic, Lawrence Katz, Robert Lucas, Derek Neal, Ariel Pakes, Jonathan Reuter, three anonymous referees, and seminar participants at the University of Chicago, the Society for Economic Dynamics annual meetings, INSEAD, New York University, the Economic Research Center, the National Bureau of Economic Research Industrial Organization meetings, Harvard University, Massachusetts Institute of Technology, the University of Chicago Graduate School of Business, the University of British Columbia, Purdue University, the University of Maryland, and the Stanford Institute for Theoretical Economics. Arie Toporovsky, Chris Xu, and Yoo-Na Youm provided excellent research assistance. We are grateful for the financial support of the National Science Foundation (SES-0242031) and the John M. Olin Foundation. Correspondence can be addressed to either of the authors at the Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637.

I. Introduction

An investor seeking to hold assets in a mutual fund is a consumer with many choices: in 2001, there were 8307 U.S. mutual funds in operation. If one counts different share classes for a common portfolio as separate options available to an investor, the implied total number of funds to choose from exceeds 13,000. Note in comparison that there were a total of 7600 companies listed that year on the NYSE, AMEX, and Nasdaq combined. A mutual fund investor's choice set has also been growing robustly over time: while there were 834 mutual funds in operation in 1980, this nearly quadrupled to 3100 by 1990, and almost tripled again by 2001.[1]

An additional, less documented feature of the mutual fund marketplace is the enormous dispersion in the fees (prices) investors pay to hold assets in funds, a dispersion that persists despite the competition among large number of industry firms. These fee differences are not simply a result of variation across fund sectors; price dispersion within (even narrowly defined) sectors is large. Table I summarizes this within-sector dispersion. The table shows fund fee dispersion moments—the coefficient of variation, the interquartile price ratio, and the ratio of the ninetieth percentile to the tenth percentile price—for each of 22 fund objective sectors in 2000.

As is evident in the table, the seventy-fifth percentile price fund in a sector-year cell typically has investor costs about twice those of the twenty-fifth percentile fund. The ninetieth-tenth percentile price ratios indicate between three- and seven-fold fee differences. The extrema of the distribution (not shown) can exhibit vast dispersion; the minimum-price aggressive growth fund, for example, imposed annualized fees of only 14 basis points (i.e., 0.14 percent of the value of an investor's assets in the fund), whereas the highest-price fund charged a whopping 1670 basis points.[2]

Of course, fund portfolios can vary considerably even within narrow asset classes. Perhaps price dispersion reflects within-sector differences in demand or cost structures across

[1] The expansion of the choice set has accompanied a steady increase in the fraction of the population taking advantage of the mutual fund option. Only 6 percent of households held mutual funds in 1980. By 2001, fully 52 percent of U.S. households held assets in mutual funds [Investment Company Institute 2002].

[2] The competitive effect of entry appears to be weak throughout the industry. Entry coincides with increases in both average fees and fee dispersion. We regress sectoral price (fee) dispersion moments for 1993-2001 on the logged number of sector funds, allowing for sector-specific intercepts and trends. The results, available from the authors, show that average fees in the sector actually increase significantly as the number of sector funds rises. The asset-weighted mean also increases (albeit insignificantly). Several fee dispersion moments (standard deviation, coefficient of variation, interquartile range, and the ninetieth/tenth and ninety-fifth/fifth percentile fee ranges) are positively correlated with increases in the number of sector funds as well.

fund portfolios. On the demand side, certain portfolios will outperform their sector cohorts; higher prices may just reflect investors' willingness to pay for better performance. As for costs, fund managers create portfolios using different securities, some of which may be more expensive to analyze or trade than others. Fund prices may reflect this fact.[3] Portfolio differentiation, too, may explain in part the large number of industry funds. Investors differ in their ideal portfolios and their current asset compositions. Perhaps thousands of funds and several hundred new funds each year are necessary to provide the many risk-return profiles sought by investors.[4]

However, a look at the retail (i.e., non-institutional) S&P 500 index fund sector strongly suggests that the composition and financial performance of funds' portfolios are not the only factors explaining fund proliferation and fee dispersion. All funds in this $164 billion (in 2000) sector explicitly seek to mimic the same performance profile, that of the S&P 500 index. Thus any discrepancies among these funds' financial characteristics should be minimal, and the observed competitive structures of the sector (including the important fund proliferation and price dispersion issues discussed above) are likely to be driven by non-portfolio effects.

It is readily apparent that, despite the sector's financial homogeneity, the features of the broader mutual fund industry are equally prominent. There were 85 retail S&P 500 index funds operating in 2000, a number that seems well beyond the saturation point arising from simple portfolio choice motives, given that each one offered conceivably equivalent expected risk-return profiles. Entry has been brisk too: the number of funds in the sector has more than quintupled since 1992. As for price dispersion, the highest-price S&P 500 index fund in 2000 imposed annualized investor fees nearly 30 times as great as those of the lowest-cost fund: 268 vs. 9.5

[3] To check the price-dispersion/performance-dispersion hypothesis, we regress gross annual returns, the within-year average gross monthly return, and the within-year standard deviation of monthly returns on fund prices. The sample consists of all mutual funds in the CRSP database between 1993 and 2001 with return data available—roughly 83,300 fund-year observations. The specifications include interacted Strategic Insight objective category (of which there are 193) and year effects, so estimated coefficients reflect the correlation between returns and prices within sector-year cells. The price coefficients in both gross return regressions are actually negative (though statistically insignificant); more expensive funds have lower-than-average returns. (A regression of net annual returns on price yields a significantly negative coefficient; one would expect zero in a perfectly competitive market where price differences are exactly compensated by gross returns.) Furthermore, the correlation between fund prices and return variance is positive and significant—also the opposite sign one would expect if performance and price were closely linked. A more careful investigation would obtain measures of expected returns and use longer performance histories to measure within-fund return variation; however, given the magnitude of the observed price dispersion, these results suggest that the price-performance link is not an overwhelming determinant of the observed patterns in the data. The findings are also in line with Carhart [1997], where the impact of expenses on performance was negative and at least one-for-one. Detailed results are available from the authors upon request.

[4] See Mamaysky and Spiegel [2001] for an argument along these lines.

basis points. Table I shows this striking divergence is not restricted to the far ends of the distribution; the seventy-fifth/twenty-fifth and ninetieth/tenth percentile price ratios are 3.1 and 8.2, respectively, which are both at the high end of the range among broader sectors. Even more interestingly, high-price funds are not all trivially small. The highest-fee fund held 1.1 percent of sector assets—enough to make it the tenth-largest fund in the sector and not much smaller than the 1.4-percent share of the lowest-price fund.[5]

This paper addresses the following puzzle: How can so many firms, charging such diffuse prices, operate in a sector where funds are financially homogeneous? We deepen the puzzle in Section II by discussing observed price heterogeneity and entry patterns among S&P 500 index funds. Section III discusses and provides empirical evidence for the existence of several factors that may explain the observed price dispersion. We provide in Section IV a model of competition in this industry that explicitly incorporates the two factors we deem to be the most important: investors' tastes for product attributes other than portfolio composition, and informational (or search) frictions that deter investors from finding the fund offering highest utility (net of management fees). Section V uses equilibrium conditions of our model to evaluate the ability of search and non-portfolio differentiation to qualitatively and quantitatively explain patterns in the data. In Section VI, we use estimated parameters of our model to quantify the social welfare implications of having so many funds delivering what are arguably ex-ante identical returns.

Estimation of our model yields the following results: product differentiation plays an important role in this "seemingly" homogeneous product industry. Investors appear to value funds' observable non-portfolio attributes, such as fund age, the total number of funds in the same fund family, and tax exposure, in largely expected ways. Our estimates also indicate that after accounting for (vertical) product differentiation across index funds, fairly small search costs can rationalize the fact that the index fund offering the highest utility does not capture the whole market. Perhaps more interestingly, our estimates imply that the distribution of search costs across investors that sustains observed prices and market shares has been shifting over time. While search costs were falling in the lower three quartiles of the distribution throughout our

[5] The sizeable price dispersion is not driven simply by loads; considerable spreads are observed among annual fees (the sum of management and 12b-1 fees) alone. The comparable dispersion measures for these fees are as follows: seventy-fifth-twenty-fifth percentile price ratio = 2.1; ninetieth-tenth percentile ratio = 6.0; and max-min ratio = 20.8.

observation period (the late 1990s), costs at the high end of the distribution were rising. We show evidence suggesting that this may be due to a compositional shift in demand: the documented influx of novice (and high-information-cost) mutual fund investors during the period, whose purchases may underlie the observed shift in sector assets to more expensive and often newly entered funds. Given our estimates of demand parameters, our welfare calculations indicate that restricting sector entry to a single fund might yield nontrivial gains from reduced search costs and productivity gains from returns to scale, but these gains may be counterbalanced by losses from monopoly market power and reduced product variety.

We see the broader contribution of this paper as twofold. The mutual fund literature is not typically concerned with strategic interactions among firms in this important industry.[6] Yet competitive forces are an important determinant of the fortunes of funds and fund families as well as consumer welfare. We seek to partially fill this research gap by providing a model of competition in this industry and taking this model to data to quantify the economic forces at work.[7] The second contribution of the paper is methodological: the modeling and estimation framework developed here can be applied to other industries where search frictions co-exist with product differentiation. Unlike previous empirical applications of models of search equilibrium (mostly in labor economics), we do not have data on individuals' decisions and must draw inferences regarding search behavior from aggregate price and quantity data.[8] Our methodology is closest in this respect to Hong and Shum [2001], who discuss identification and estimation of equilibrium search models using only market-level price data from markets for homogeneous goods. Our model extends their results by showing how aggregate price and quantity data can be used to identify and estimate search costs separately from sources of vertical product

---

6 This literature, too vast to cite comprehensively, follows from the classic contributions on portfolio theory and empirical testing as well as the principal-agent framework. See, for example, Jensen [1968], Malkiel [1995], Falkenstein [1996], Gruber [1996], and Chevalier and Ellison [1997, 1999].

7 We should note that there has been a recent increase of interest among financial economists in this area. For example, Sirri and Tufano [1998] note that mutual fund flows are relatively insensitive to management fees and excessively sensitive to past performance (as opposed to expected performance). Khorana and Servaes [1999, 2003] empirically explore what market characteristics are correlated with fund entry and the market share of fund families. Barber, Odean, and Zheng [2002] find that fund inflows seem to be more responsive to some price instruments than others, and that advertising appears to be an effective tool for increasing fund assets. On the theoretical side, Massa [2000] and Mamaysky and Spiegel [2001] explore the driving forces of fund creation.

8 Sorensen [2001] is a recent attempt in the industrial organization literature to estimate search costs using consumer-level product choice data. Examples of structural estimation of search models in labor markets include Flinn and Heckman [1982], Eckstein and Wolpin [1990], and van den Berg and Ridder [1998].

differentiation, with particular attention to minimizing the impact of functional form restrictions.

II. An Overview of the Retail S&P 500 Index Funds Sector

S&P 500 index funds are the most popular type of index mutual funds. Their explicit investment objective is to replicate the return patterns of the S&P 500 index. Many do so by holding equities in the same proportions as their index weights, while some use other index-matching methods like statistical sampling and index derivatives.

Despite the existence of different index replication strategies, there is considerable financial performance homogeneity in our sample of S&P 500 index funds.[9] The top rows of Table II contain summary statistics of our sample funds' gross annual returns and standard deviations of their monthly gross returns. As can be seen, the relative dispersion in return patterns is quite small. The interquartile range of gross annual returns is no greater than 0.32 percentage points, and is typically about 1 percent of the mean.[10] The dispersion of the standard deviation of funds' monthly returns is similarly slight: the average interquartile range-mean ratio is 0.007, and the coefficient of variation is under 5 percent in every year. These small variations suggest investors would be justified in presuming common ex-ante returns among our funds.[11]

Despite this homogeneity in financial performance, the sector shares the fund proliferation and price dispersion traits seen in the broader mutual fund industry. There was vigorous growth in both the number of retail S&P 500 index funds and their total net assets under management from 1995 to 2000, our period of observation. This was coincident with and

[9] We exclude institutional S&P 500 funds from our sample, despite the fact that they also mimic the same performance profile, because we believe they operate in a fundamentally different product market than non-institutional funds (more on this below). By doing so, we hope to further control for unobservable differences across funds that might confound our analysis. We also exclude "enhanced" index funds, which blend active trading strategies with passive index-holding.

[10] Statistics on returns in the table correspond only to funds operating every month in observed year to eliminate composition bias from funds with return data spanning only a possibly non-representative portion of the year. While the dispersion patterns are largely mirrored in the standard deviations of the return moments, outliers do enlarge the standard deviation of annual returns in some years, to as much as 14 percent of the mean in 2000. An index fund's performance can deviate from the index which it is tracking (and from other funds tracking the same index) because of several factors. These include idiosyncratic portfolio sales required to meet the particular daily activity needs of a fund, how much of the fund's assets are held in cash, and the timing of trades.

[11] As with the broader sample discussed above, a regression using our S&P 500 funds indicates a significantly negative correlation between net annual returns and price, suggesting higher prices are not being compensated by higher gross returns of equal size. Also as with the broader sample, a regression of the standard deviation of funds' monthly returns on price indicates a positive and significant correlation. This is of course the opposite sign predicted by a close price-performance link. (Year fixed effects are included in both regressions.)

even stronger than the overall growth in the mutual fund industry documented above. The number of sector funds more than tripled from 1995 to 2000 (from 24 to 85), and sector assets grew at twice the rate as did total equity fund assets.[12]

Figure I and the middle portion of Table II show the evolution of the sector's price distribution over our sample period. A striking feature of Figure I, which plots the cumulative price distribution functions of sector funds, is the rightward shift in the price distribution from 1996 to 1999. (This trend interestingly reversed in 2000.) The movement was steady; the 1997 and 1998 distributions mark continuities in the evolution. The shift occurred throughout the distribution, but particularly at the high end. As can be seen in Table II, this shift happened despite the entry of nearly a dozen funds a year and a steady drop in concentration during the period. Along with the increase in average price evident in the figure, the table also documents an increase in price dispersion: the interquartile range and standard deviation of prices increase (though not monotonically) over the observation period.

An interesting related observation is the change in the relative market shares of the high- and low-price segments. It is evident in the bottom of Table II that while low price funds still dominate the market, the asset share of funds in the lowest price decile has fallen consistently since 1995. In contrast, the market shares of the upper quartile (and decile, not shown) rose during the same period. The proportional rise in sector assets held in expensive funds has been especially stark; the market share of the top quartile nearly tripled. The reallocation of market share to higher-priced funds resulted in a 20 percent rise in the sector's asset-weighted average price from 1995 to 2000, from 26.8 to 32.2 basis points. Indeed, if the asset-weighted mean price had held constant at 1995 levels, total annualized fees would have been $88.5 million lower in 2000.

Recall that these increases in price dispersion and levels are occurring in concert with robust entry into the sector. An examination of entry patterns sheds additional light on this issue. From 1995 through 1999, while 25 funds entered the sector with prices above 100 basis points, only seven funds charging less than 40 basis points entered. Using the decomposition method of Foster, Haltiwanger, and Krizan [2002], we found that the entrants' asset-weighted average price

[12] Sector growth was driven in part by the growing popularity of passive management strategies among investors. The S&P 500 index fund sector was particularly able to capitalize on this preference shift partly because the original index fund, the First Index Investment Trust, is today the Vanguard 500 Index Fund. Its early entry into a growing sector has been seemingly very important: the Vanguard 500 Index Fund still is a dominant player in the sector.

was higher than the weighted mean among incumbents in each year from 1995 to 1999. Moreover, entry was the predominant contributory factor in the two years with the largest average share-weighted price increases (1998 and 1999).[13]

III. Sources of Price Dispersion

We have documented above that portfolio differentiation is an unlikely explanation of the observed price dispersion and entry/asset-reallocation patterns. In this section, we discuss and provide empirical evidence for and against three mechanisms that might explain these seeming anomalies: non-portfolio "product" differentiation of mutual funds, search costs/information frictions, and switching costs. This discussion will motivate our model of industry equilibrium and its detailed empirical investigation in Sections IV, V, and VI.

III.A. Non-Portfolio Differentiation

Fund attributes other than portfolio composition are one possible reason investors would value funds differently. If portfolio returns come bundled with a set of services that differ across funds, price dispersion among financially homogeneous funds could be sustained. Capon, Fitzsimmons and Prince [1996] report from a survey administered to roughly 3400 mutual fund holders that investors consider the number of other funds in the family and the fund company's responsiveness to enquiries as the third- and fourth-most important criteria (respectively, after performance and manager reputation) in their mutual fund purchase decision. These fund attributes are not directly related to the financial characteristics of a fund's portfolio, and could well account for price differences across funds with equivalent return profiles.

Other such attributes might play a similar role. For example, the provision of financial advice, usually bundled with the purchase of load funds, is important to many investors: 60 percent reported consulting a financial adviser before purchase (Investment Company Institute 1997). We explore this influence in detail below. Other potentially important attributes explored below are whether the fund is an exchange-traded fund, its age, the tenure of its manager, its rate of taxable distributions, and the quality of its account services (frequency and quality of account statements, account access by phone, etc.).

[13] See Table 4 of Hortaçsu and Syverson [2003], which breaks changes in the (asset-based) market-share-weighted mean price into within, between, and net entry components (also see ibid. footnote 16).

We incorporate such non-portfolio differentiation into our model in Section IV. Funds differ in their offered utility levels because they have varying amounts of attributes valued by investors. We assume all investors place the same values on these qualities. It seems possible, though, that horizontal taste differences (like differing tastes for certain attributes or a logit-type random utility term) may exist in conjunction with the modeled vertical component. We have explored more general preference specifications that incorporate both horizontal taste differences and search cost variation across consumers. However, as will be seen below, market outcomes driven by horizontal differentiation cannot in general be separately identified in our data from those caused by search costs. This unfortunately prevents us from estimating such models and directly testing for horizontal taste variation. We do, however, estimate a specification that allows a particular form of horizontal differentiation where investors differ by type in terms of whether they buy a fund through an intermediary bundling advisory services for a load (sales charge), or through "direct" channels that lead to no-load fund purchases. As will be discussed, this specification reflects the institutional setup of the industry well, and provides interesting insights into structural changes in demand.

III.B. Search Costs/Information Frictions

An additional (but not mutually exclusive from product differentiation) possible explanation for the observed price dispersion is the influence of search/information frictions faced by investors. A large theoretical literature shows that costly search can sustain price dispersion in homogenous product markets [e.g., Burdett and Judd 1983, Carlson and McAfee 1983, Stahl 1989]. Given the very large number of mutual funds offered, it seems reasonable to presume that investors must make some information-gathering investments before deciding between fund alternatives. The presence of a sizeable market to reduce investor search costs supports this notion. Several commercial mutual fund ranking services and information aggregators exist (Morningstar, Lipper, Valueline, Yahoo!Finance, etc.). There is even a commercial Internet site (IndexFunds.com) devoted to providing information about index funds. Many fund companies spend considerable sums on marketing and distribution, also consistent with (although neither necessary nor sufficient for) the presence of limited investor information. Survey evidence also suggests considerable information-gathering. The Investment Company Institute [1997] reports that surveyed investors consulted a median of two source types (four for those who had consulted a fund-ranking service) and reviewed a median of 14 different information items (gross returns, relative performance, etc.) before their most recent purchase.[14] To the extent that gathering and analyzing such information consumes investor time and money, these activities constitute costly search.

Aside from this anecdotal evidence, price patterns among institutional-class S&P 500 index funds offer more direct substantiation of the importance of search frictions in the sector. As mentioned before, we do not include these institutional funds in our analysis because of our belief that they operate in a fundamentally different product market. The very high initial minimum investment levels (typically at least $1 million) restrict demand to a fairly narrow class of investors. It also implies that if there is investor search, there is a larger gain (in absolute levels) to finding a lower-price and/or higher-quality fund for a typical institutional buyer than for a retail investor. It might then be reasonable to assume there are higher search intensities among institutional funds, implying less price dispersion and lower average prices. The data are consistent with this. Figure II compares histograms of institutional and retail fund prices in 2000. It is readily apparent that the former distribution is considerably tighter and has a smaller mean. While administrative cost advantages may be in part responsible for the lower average price of institutional funds, they are unlikely to affect price *dispersion*. We cannot rule out all other explanations for these differences, but this prima facie evidence is suggestive of search.[15]

III.C. Switching Costs

An additional explanation for some of the patterns documented above are the switching costs involved when investors move assets across fund families (intra-family transfers are typically free). These costs could either be "formal" (like those created by deferred or rear loads) or "informal" (the hassle associated with drafting a letter to the fund company to approve withdrawals, for example). While in many ways similar to search costs, switching costs are

[14] The survey was administered at the very beginning of Internet access diffusion among the general public. The use of online sources has surely risen since then, raising the possibility that per-fund search costs have declined since the beginning of our sample period. We address issue this in more depth below.

[15] In particular, this evidence (as well as some of the other patterns described above, perhaps) is also consistent with Gabaix and Laibson's [2003] intriguing suggestion that mutual fund companies "confuse" boundedly rational investors by giving them noisy signals about the surplus from their fund purchase—e.g., by utilizing complicated nonlinear pricing schemes. They show that such endogenously supplied "confusion" enables price dispersion to be sustained among homogeneous products and results in entry leading to price increases. Our evidence is consistent with their explanation in that institutional investors are likely to be less bounded in their rationality and hence more difficult to "confuse" than retail investors—sustaining lower levels of price dispersion.

distinct. With search, an investor considering moving assets to another family knows the utility benefit from staying but must pay a cost to learn the benefit of going elsewhere. Switching costs, on the other hand, imply that the investor knows the utility benefit of moving her money to another fund company, but a cost must be paid to move. To separately identify these two factors, we would need to exploit variation in investors' information sets. Unfortunately, this decomposition is very difficult if not impossible without detailed investor-level data, so we cannot directly test a "search vs. switching costs" hypothesis in our model.

Still, some information is available that allows us to obtain a general idea of the size of switching cost effects, and we attempt to do so here. Informal switching costs are difficult to quantify, but the mechanism by which formal switching costs like rear and deferred loads operate is obvious: investors with rear/deferred-load fund holdings must pay a charge (typically up to 5 percent) if they remove those assets from the family within a specified time period. For such investors, this gives the S&P 500 index fund run by the load fund family an advantage over (often lower-priced) S&P 500 index funds offered by others.[16] In particular, such load-driven switching costs may lead to "parking" behavior, where investors holding non-S&P 500 index funds in the family move assets over to (or "park" in) the family's S&P 500 fund when dissatisfied with the performance of the other (possibly actively managed) funds in the family. Given that several fund families specializing mostly in actively managed funds opened an S&P 500 index fund in this period, one might conjecture that the market share gains made by these entrants (typically high-priced, as we have shown) is driven in part by the captive demand of "parkers."

To gauge the empirical importance of such behavior, we search for patterns that one might expect to see in our data if switching-cost-induced parking behavior is important. One pattern regards the response of asset flows into our S&P 500 index funds to the relative performance of other funds within the same fund family. It seems likely that the overall performance of a family's funds influences flows into or out of specific funds within the family, say through "star" fund effects or similar performance-spillover mechanisms. If such a performance spillover exists, high (low) average performance of a fund family's non-S&P 500

[16] Front loads, while sunk, still create switching costs through a more subtle mechanism. If front load fund owners remove assets from the family, they lose the option value of being able to transfer those assets among that family's funds without paying additional loads. Once the assets are moved out, another load must be paid should they choose to bring them back in. (We thank a referee for pointing out to us this option value.)

funds should, all else being equal, boost (decrease) net asset inflows into the family's S&P 500 fund. However, this spillover into the S&P 500 fund should be damped in load fund families if switching-cost driven parking is important: when a load fund family's non-S&P 500 funds experience a period of low (high) performance, the parking response will cause investors holding assets in the family's other funds to move money into (out of) the family's S&P fund. This movement would counteract in part the spillover-driven flows.

To test this hypothesis, we regress annual net dollar flows into our S&P 500 index funds on an asset-weighted average performance measure of the non-S&P 500 index funds in the same family. The performance measure we use for the non-index funds is the weighted mean of the Sharpe ratio of these funds' excess returns relative to the S&P 500 fund (or a weighted average among the index funds if there is more than one in a family). After including year dummies and fund fixed effects in our regressions, we find a positive and significant correlation between flows into the S&P 500 fund and the performance of the non-S&P 500 funds in the family. This finding is consistent with our presumption that performance spillovers exist. However, when we allow the magnitude of the response of S&P 500 index fund flows to the performance of other funds in the family to differ between load fund families and no-load fund families, we do not find a statistically significant difference. This goes against the hypothesis that switching costs in load fund families damp the spillover-induced net flows to their S&P 500 index funds.

A second empirical pattern one may expect to observe if switching costs and parking behavior are important is that *within-family* asset share of S&P 500 index funds should grow faster in load fund families than in no-load families. Sector growth over our observation period led to virtually all S&P 500 funds comprising larger shares of their family's overall assets over time. If switching costs are important, the introduction of a new S&P 500 index fund in a load fund family may induce captive within-family investors to switch into this new offering, in addition to any new investors arriving at the fund family. This suggests that the within-family asset share of S&P 500 index funds in load fund families may grow faster than their counterparts in no-load families, as fund holders in load families take advantage of their newfound ability to park their assets in the index fund.

We do not find evidence of this second pattern in our data. We regress (at the fund family level) the change in the within-family asset share of their S&P 500 index funds on a set of year dummies and an indicator for load families. The estimated coefficient on this latter dummy is negative and insignificant. A similar regression was run with the growth rate of the within-family share (rather than just the share change) as the dependent variable. Here the load-family dummy was positive and insignificant. Hence, rather than the faster growth in load fund families implied by high switching costs, it seems that there is no appreciable difference across load and no-load families in the growth of within-family asset shares of S&P 500 index funds.

While we do not interpret the outcomes of these simple tests as evidence that *no* load-driven switching costs exist in the sector, they do suggest to us that their effects might not be as large as one may think. We also note that the concept of a typical investor locked into holding assets in a single fund family due to switching costs is inaccurate: for example, the Investment Company Institute [2000] reports the median number of mutual fund companies with which investors hold assets is two.

Given this evidence, the theoretical model we will develop in the next section will not include switching costs as an explicit component. Still, we believe that switching costs created by the fund family structure of the industry play some role in the S&P 500 index fund sector. We attempt to account for this influence in our empirical model—albeit in an imperfect manner—by letting investors' purchase decisions depend on the size of funds' management companies, and by allowing investors to respond to the presence of a rear/deferred load independently from the presence of bundled advice. Furthermore, given the conceptual similarity between switching and search costs discussed above, we expect that some switching cost effects are likely to be reflected in our estimates, suggesting a broader interpretation of our estimates of "investor search costs" beyond information-gathering costs alone.

IV. The Model

IV.A. General Setup

Demand for sector funds is characterized by a continuum of investors searching over funds with varying attributes.[17] Investors have heterogeneous search costs, and we also allow for the fact that some index funds are "easier to find" than others by allowing heterogeneous sampling probabilities across funds. We assume that fund attributes are vertical characteristics, and that all investors share a common utility function. Thus conditional on investing in fund j, an investor receives indirect utility equal to $u_j(W_j;\ \theta)$, where W_j is a vector of fund attributes and

[17] See Anderson and Renault [1999] for an alternative model with costly search over differentiated products.

θ is a set of parameters that characterize how the attributes affect utility. The model requires few specific assumptions on the form of $u_j(\cdot)$ a priori. However, for reasons that will become clear in the next section, we assume utility is a linear function of fund characteristics:

$$u_j = \underline{W}_j \beta - p_j + \xi_j, \tag{1}$$

where $\underline{W}_j$ are the elements of W_j other than price p_j and an unobservable component ξ_j. Note that the coefficient on the price term has been normalized to –1, so utilities are expressed in terms of the unit of price measurement. Here, given the nature of the good, that unit is basis points. Thus one can think of u_j as specifying fund utility per dollar of assets the investor holds in it.

Despite a clear rank ordering of funds, the fund delivering the largest u_j does not gain 100 percent market share because search is costly. We assume search costs are heterogeneous in the investor population and have distribution $G(c)$. Investors incur this cost to learn the indirect utility of a particular fund, with the exception of the first fund they search (this assures all investors desiring to hold assets in the sector end up doing so regardless of their search cost level). For tractability, we assume that investors search with replacement, and are allowed to "revisit" previously searched funds. We also restrict investors to only purchase shares in one S&P 500 index fund.[18]

Define investors' belief about the distribution of funds' indirect utilities $H(u)$. Then the optimal search rule for an investor with search cost c_i is to search for another fund as long as

$$c_i \leq \int_{u^*}^{\bar{u}} (u - u^*) dH(u), \tag{2}$$

where $\bar{u}$ is the upper bound of $H(u)$, and u^* is the indirect utility of the highest-utility fund searched to that point. This is a standard condition in sequential search models; search continues if the marginal cost of search is no greater than the expected marginal benefit. We simplify matters by assuming that investors observe the empirical cumulative distribution function of funds' utilities. That is, label the N funds by ascending indirect utility order, $u_1 < \ldots < u_N$. Then

[18] Following Carlson and McAfee [1983], the search-with-replacement assumption greatly simplifies matters in search models involving a finite number (as opposed to a continuum) of products because we do not need to worry about how investors' beliefs about the price distribution evolve as certain funds are removed from consideration. This deviation from reality is small when there are a large number of funds. The revisit assumption implies of course that the investor's benefit from searching is relative to the best fund yet searched, rather than the particular fund in hand at any given time.

(3) $$H(u) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{I}\left[u_j \le u\right].$$

Thus investors know the available array of indirect utilities; they just do not know which fund provides what utility level until engaging in costly search.

The optimal search rule yields critical cutoff points in the search distribution, given by:

(4) $$c_j = \sum_{k=j}^{N} \rho_k \left(u_k - u_j\right),$$

where ρ_k is the probability that fund k is sampled on each search (these probabilities are known by investors), and c_j is the lowest possible search cost of any investor who purchases fund j in equilibrium. The intuition behind this expression is as follows. Optimal search continues until the investor's expected benefit from searching is lower than the search cost. The right-hand side of expression (4) is the expected benefit of additional search for an investor who has already found fund j. This is product of the probability ρ_k that another search yields a higher-utility fund (recall $u_k > u_j$ if $k > j$) and the corresponding utility gain $u_k - u_j$, summed over all funds superior to fund j. Fund draws with utilities less than u_j are ignored in this calculation, as investors can costlessly revisit funds already searched. Note that this expected benefit declines in the fund's index; in fact, $c_N = 0$. Thus as long as an investor's search cost is lower than c_j, he or she keeps searching until a fund offering greater utility than fund j is found. On the other hand, more search is not worthwhile for any investor with search costs greater than c_j who finds fund j. Notice that this implies the product index is declining in the ordinal ranking of critical search cost values; i.e., while $u_1 <\ldots< u_N$, $c_N <\ldots< c_1$.

We can use this optimal search behavior to identify elements of the search cost distribution. Funds' market shares can be written in terms of the search cost c.d.f. by using the search-cost cutoffs above. Consider the lowest-utility fund, u_1. This fund has a high cutoff search value, c_1, because any investor having already found this fund has a large expected benefit from continuing to search. Therefore only investors with very high search costs ($c > c_1$) purchase the fund. At the same time though, not *all* investors with $c > c_1$ purchase the fund, only those (unfortunate ones) who happen to draw fund 1 first—which happens with probability ρ_1. Thus the market share of the lowest-utility fund is

(5) $$q_1 = \rho_1\left(1 - G(c_1)\right) = \rho_1\left(1 - G\left(\sum_{k=1}^{N} \rho_k \left(u_k - u_1\right)\right)\right).$$

Now consider the market share of the second-lowest utility fund, fund 2. Again a fraction of the highest-search-cost investors ($c > c_1$), unable to afford a second search, find fund 2 first and purchase it. But a subset of investors with search costs $c_1 < c < c_2$ also purchase fund 2; namely, those who find fund 2 on their first search, or those search only to find fund 1 and keep searching until they draw fund 2. This happens with probability $\rho_2/(1-\rho_1)$.[19] Thus the total market share of fund 2 is

$$q_2 = \rho_2\left(1 - G(c_1)\right) + \frac{\rho_2}{1-\rho_1}\left[G(c_1) - G(c_2)\right] = \rho_2\left[1 + \frac{\rho_1 G(c_1)}{1-\rho_1} - \frac{G(c_2)}{1-\rho_1}\right]. \tag{6}$$

Analogous calculations, detailed in the appendix, produce a generalized market share equation for funds 3 through N:

$$q_j = \rho_j \left[\begin{array}{l} 1 + \dfrac{\rho_1 G(c_1)}{1-\rho_1} + \dfrac{\rho_2 G(c_2)}{(1-\rho_1)(1-\rho_1-\rho_2)} + \displaystyle\sum_{k=3}^{j-1} \dfrac{\rho_k G(c_k)}{(1-\rho_1-\ldots-\rho_{k-1})(1-\rho_1-\ldots-\rho_k)} \\ -\dfrac{G(c_j)}{(1-\rho_1-\ldots-\rho_{j-1})} \end{array}\right]. \tag{7}$$

These equations form a system of linear equations linking $G(c_1),\ldots,G(c_{N-1})$—the population fractions with search costs less than the distribution's critical values—to observed market shares. Moreover, we know $G(c_N) = 0$, because (4) implies $c_N = 0$ and search costs cannot be negative.[20]

The market supply side is comprised of funds that choose prices to maximize current profits. Let S be the total size of the market, p_j and mc_j be the price and (constant) marginal costs for fund j, and q_j be the fund j's market share given the price and characteristics of all sector funds. Then the profits of fund j are

$$\Pi_k = Sq_j(p,\underline{W})(p_j - mc_j). \tag{8}$$

Profit maximization implies the standard first-order condition for p_j:[21]

$$q_j(p,\underline{W}) + (p_j - mc_j)\frac{\partial q_j(p,\underline{W})}{\partial p_j} = 0. \tag{9}$$

---

[19] The total probability that a search sequence yields only fund 1 draws until a fund 2 draw is $\rho_1\rho_2 + \rho_1\rho_1\rho_2 + \rho_1\rho_1\rho_1\rho_2 + \ldots = \rho_1\rho_2/(1-\rho_1)$. Summing with the probability that the first draw is fund 2 (ρ_2) yields $\rho_2/(1-\rho_1)$.

[20] Notice that since $G(c_N) = 0$, the market share equations only include N-1 values of $G(c)$.

[21] Our search model admits pure strategy equilibria in funds' pricing decisions, rather than relying on mixed strategies to produce the dispersion. As Carlson and McAfee [1983] show, pure strategies can be supported when both consumer search costs and production technologies are heterogeneous. This is the case in our model, where funds have differing marginal costs.

The elasticities faced by the fund are determined by the derivatives of the share equations (7). We show in the appendix that these derivatives are

$$\frac{\partial q_j}{\partial p_j} = -\frac{\rho_1 \rho_j^2 g(c_1)}{1-\rho_1} - \frac{\rho_2 \rho_j^2 g(c_2)}{(1-\rho_1)(1-\rho_1-\rho_2)} - \sum_{k=3}^{j-1} \frac{\rho_k \rho_j^2 g(c_k)}{(1-\rho_1-\ldots-\rho_{k-1})(1-\rho_1-\ldots-\rho_k)}$$

$$- \frac{\rho_j \left(\sum_{k=j+1}^{N} \rho_k \right) g(c_j)}{(1-\rho_1-\ldots-\rho_{j-1})} \qquad (10)$$

Note that search cost distribution densities $g(c)$, evaluated at the cutoff values for funds offering lower utility than j (i.e., $k < j$), affect fund j's demand elasticity. To see why, consider investors' reactions to an increase in the price of fund j. The price hike decreases u_j. This has two distinct effects on the critical search cost cutoff values. For $k < j$, c_k decreases (see (4)); if you hold a fund of lower quality than j, additional search becomes less appealing when u_j declines. Thus some investors with search costs greater than c_{j-1} who would have formerly continued searching and serendipitously found fund j no longer do so. Fund j's sales lost through this channel are directly related to the density of the investor population at these higher search costs, as embodied in the first j-1 terms of (10). The second, more direct quantity effect of a price rise is from the increase in c_j when u_j falls. That is, continued search becomes more beneficial for investors who would have purchased j at the original price. Some marginal investors, the population of which is given by $g(c_j)$, now choose to continue searching and end up buying higher utility funds than j. The final term in (10) captures this loss.

This link between fund prices and the p.d.f. of the search cost distribution, as well as the connections between market shares and the distribution's c.d.f. shown above, play an important role in empirically identifying the model. We discuss this below.

IV.B. Identification

The market share equations (7) show how we can map from observed market shares to the c.d.f. of the search cost distribution evaluated at the critical values. If we know (or assume) the sampling probabilities ρ_j, then using observed market shares to solve the linear system (7) for $G(c_1),\ldots,G(c_{N-1})$ and using the fact that $G(c_N) = 0$ gives all critical values of the c.d.f. If the sampling probabilities are unknown, and must be estimated, the probabilities as well as the search cost distribution can be parameterized as $\rho(\omega_1)$ and $G(c;\omega_1)$, respectively. Given ω_1 and

ω_2 of small enough dimension, observed market shares can be used to estimate these parameters.

While market share data can be mapped into the c.d.f. of the search cost distribution, the distribution itself cannot in general be traced out using only share information. This is because market shares do not generically identify the *level* of the critical search cost values $c_1,\ldots,c_N$, only their relative positions in the distribution. However, shares *do* identify search cost levels in the special but often-analyzed case of homogeneous (in all attributes but price) products with unit demands; i.e., when $u_j = u' - p_j$, where u' is the common indirect utility delivered by the funds. In this case, (4) implies

$$(11) \qquad c_j = \sum_{k=j}^{N} \rho_k \left(u' - p_k - \left(u' - p_j\right)\right) = \sum_{k=j}^{N} \rho_k \left(p_j - p_k\right).$$

Now, given sampling probabilities (either known or parametrically estimated), $c_1,\ldots,c_{N-1}$ can be calculated directly from observed fund prices using (11).

In the more general case where products differ in attributes other than price alone, additional information is required to identify cutoff search cost values. We find this information in fund companies' optimal pricing decisions. The logic of our approach is straightforward. We need to recover the p.d.f. of the search cost distribution (evaluated at the cutoff points). These values enter the derivatives of the market share equations with respect to price, (10). If we assume Bertrand-Nash competition, the first order conditions for prices (9) imply:

$$(12) \qquad \frac{\partial q_j(p)}{\partial p_j} = -\frac{q_j(p)}{p_j - mc_j}.$$

We observe prices and market shares in the data. Therefore, given knowledge of marginal costs mc_j, we can compute $\partial q_j / \partial p_j$ using (12). From (10), these derivatives form a system of $N - 1$ linear equations that can be used to recover the values of the search cost density function $g(c)$ at the critical values $c_1,\ldots,c_{N-1}$. If marginal costs are not known, they can be parameterized along with the search cost distribution and estimated from the price and market share data.

Once the values of the search cost c.d.f. and p.d.f. (evaluated at the cutoff search costs) have been identified, we can recover the level of these cutoff search costs c_j in the general case of heterogeneous products. To do so, we note that by definition the difference between the c.d.f. evaluated at two points is the integral of the p.d.f. over that span of search costs. This difference can be approximated using the trapezoid method; i.e.,

$$(13) \qquad G\left(c_{j-1}\right) - G\left(c_j\right) = 0.5\left[g\left(c_{j-1}\right) + g\left(c_j\right)\right]\left(c_{j-1} - c_j\right).$$

We invert this equation to express the differences between critical search cost values in terms of the c.d.f. and p.d.f. evaluated at those points:

$$(14) \qquad c_{j-1} - c_j = \frac{2\left[G(c_{j-1}) - G(c_j)\right]}{g(c_{j-1}) + g(c_j)}.$$

Given the critical values of $G(c)$ and $g(c)$ obtained from the data above, we can recover the c_j, and from these trace out the search cost distribution.[22] In non-parametric specifications, a normalization is required: the demand elasticity equations do not identify $g(c_N)$, so a value must be chosen for the density at zero-search costs (recall that $c_N = 0$).[23]

Furthermore, we can also use the critical values of the search cost distribution c_j to estimate the attribute loadings β in the utility function (1). Funds' implied indirect utilities u_j are calculated from the c_j obtained above using the linear system (4).[24] Since we impose a price coefficient in the utility function of –1, we can then estimate β with the following regression:

$$(15) \qquad u_j + p_j = X_j \beta + \beta_{age} \ln(age_j) + \eta_j,$$

where X_j are observed fund attributes other than age, and η_j is a fund-specific error term. Because the unobservable fund attribute ξ_j, which is included in η_j, is likely to be correlated with fund age (survival/longevity should be positively related to unobservable quality because high-ξ_j funds are less likely to exit), we treat fund logged age as endogenous and estimate (15) using instrumental variables. We use the instruments used by Berry, Levinsohn, and Pakes [1995] to estimate utility functions with unobservable quality. These are current-year own-fund attributes (to instrument for themselves) and current-year summary measures (fund counts and average levels of each non-age observable attribute) of two other sets of sector funds: other funds managed by the same company (excluding the own-fund observation), if applicable; and those managed by other companies. These two sets are unique to each fund, so instruments vary by

[22] Of course, since we only identify the search cost distribution at the cutoff values c_j, we don't identify the c.d.f through its entire domain. Any monotonically increasing function between the identified cutoff points could be consistent with the true distribution; our trapezoid approximation essentially assumes this is linear. The approximated c.d.f. converges to the true function as the number of funds increases.

[23] The intuition for this is that demand elasticities are determined by the actions of searchers on the margin between two funds. Given that search is responsible for spreading output shares across funds, and that changes in indirect utilities move shares on the margin only between adjacent funds, there are only N-1 margins for N funds. Thus the markup/elasticity equation system only identifies the first N-1 cutoff values of the search cost density function.

[24] Note that in our current setup, (4) implies that $u_1 = 0$, so fund utility levels are expressed relative to the least desirable fund. This normalization results from our assumption that all investors purchase a fund. If we added an outside good that could be purchased without search, we could alternatively normalize this good's utility to zero.

fund-year observation. These instruments are meant to capture the effect that a fund's relative position in attribute space (which is assumed exogenous) has on the exit decision, independent of the fund's unobservable quality.

It is interesting to note the links between our demand system and that implied by a standard discrete choice demand model like the multinomial logit [e.g., McFadden 1974]. Ours has purely vertically differentiated products, but still implies a nondegenerate market share distribution because search cost variation across investors creates a type of horizontal differentiation. The standard logit model also introduces products that are (almost) purely vertically differentiated, but builds in horizontal differentiation (and its resulting market share dispersion) directly into the preference function with the i.i.d. random utility term.

One may notice that we do not explicitly build horizontal *taste* differentiation into the model. However, in section V.E, we consider an extension that allows the investor population to be horizontally differentiated according to their mutual fund purchase channel. Further, we could in principle also build in horizontal differentiation into our model by allowing investors' tastes for fund attributes to be drawn from a parametrizable distribution, as in Berry, Levinsohn and Pakes [1995]. Identification of the taste distributions' parameters would require across-market variation in the choice sets faced by investors. However, to empirically separate taste heterogeneity from search cost differences in our model, we would also need to observe something that moves the investor search cost distribution *independently* of tastes. Unfortunately, our data does not offer a source of such variation, and as such we cannot directly test for the presence of general forms of horizontal taste variation.

V. Estimation

V.A. The Basic Model

Our approach to estimating the model is to build up from the simplest version of the model, adding complexities (sometimes at the cost of parametric assumptions) as we go along, and compare the performance of the various versions in explaining the data.

We begin by assuming that funds are homogeneous: the only characteristic that matters to S&P 500 index fund investors is price. As noted above, homogeneity implies that $u_j = u' - p_j$, where u' is common to all funds, and given sampling probabilities, the cutoff search cost values can be computed directly from observed prices.

Consider the case where funds have equal sampling probabilities; i.e., $\rho_j = 1/N \; \forall \; j$. In this case the market share equations (7) simplify to:

$$(16) \qquad q_j = \frac{1}{N} + \sum_{k=1}^{j-1} \frac{1}{(N-k+1)(N-k)} G(c_k) - \frac{1}{(N-j+1)} G(c_j).$$

As noted above, this system non-parametrically identifies the c.d.f. of the search cost distribution. The implied $G(c_j)$ values, when combined with the computed c_j from the version of (4) with $\rho_j = 1/N \; \forall \; j$ and $u_j = u' - p_j$, would allow us to trace out the search cost distribution.

However, this simple version of the model is rejected straightaway by the data. To see why, note that when sampling probabilities are equal and funds homogeneous, there will be a negative and monotonic relationship between price and market share. A glance at Figure III, which plots the fund price vs. market share (both in logs) for the 2000 funds, shows that this is not the case in the sector.[25] While there is a clear negative correlation between price and market share, the relationship is far from monotonic. (Recall as an example of this departure from monotonicity that the highest-fee fund was the tenth-largest of the over 85 funds in 2000.)

The rejection of this simplest version of the model indicates that fund differentiation must matter on some level. We consider two possibilities. One is that funds are perceived by investors as homogeneous, but the likelihood of "finding" a fund is a function of the fund's attributes. This breaks the basic model's implication of monotonicity between fund price and market share by letting certain higher-priced funds be sampled with a higher relative probability than some of their lower-priced competitors. The second possibility, already alluded to above, is that investors view funds as differentiated in non-price characteristics. A higher-priced fund may then have a larger market share than its cheaper competitor because it provides other attributes investors value. We consider both of these more complex versions in turn.

V.B. Unequal Sampling Probabilities

To allow sampling probabilities ρ_j to vary across funds, we use the following functional form specification:

$$(17) \qquad \rho_j = \frac{Z_j^{\alpha}}{\sum_{k=1}^{N} Z_k^{\alpha}},$$

[25] Other years show a similar pattern; we choose 2000 because it had the largest number of funds.

where Z_j is an index of fund level observables that influence the probability that the fund sampled. The parameter α captures any nonlinearities. What should enter into Z_j? If sampling probabilities do differ, they are likely correlated with a fund's marketplace exposure. Since the modeled search process stylizes a very complex process where investors learn about and choose among alternative funds, it may be reasonable to assume certain funds have more visibility than others and are thus more likely to be considered by investors. One possible exposure measure is advertising expenditures. Unfortunately, we do not observe this at the fund level. Reasonable proxies would be variables related to (but not completely collinear with) the size of the fund, to capture the dynamics of the search process and possible "social learning" effects. We use fund age to embody these influences in our benchmark Z_j and then consider extensions below.

We estimate the search cost distribution and sampling probabilities by fitting the market share equations (7) and the first-order pricing conditions (9) using nonlinear least squares. The sample consists of all retail S&P 500 funds operating in any year between 1995 and 2000 (inclusive), excluding ten fund-year observations for which we do not have fee or asset data. This leaves a sample of 309 fund-year observations. Search costs are parameterized as lognormal with $E[\ln(c)] = \mu$ and $\mathrm{Var}[\ln(c)] = \sigma^2$. We add further flexibility by allowing both of these moments to have trends over the sample. The mean marginal cost and the sampling probability parameter(s)—assumed constant throughout the sample—are also estimated.[26]

The results are presented in Table III. Five slightly different versions are estimated. The benchmark model, in column (A), uses the asset-based market shares and the literature-standard seven-year annualized prices upon which the empirical work above is based. Column (B) shows the estimates from as specification that uses market shares measured using money flows into funds rather than asset stocks.[27] The two versions (C) and (D) explicitly account for possible price incidence differences between load and no-load funds. While the annualized prices of no-

[26] A lognormal specification fit the data much better than normal, exponential, uniform, and gamma distributions. We show below that a lognormal distribution also fits nonparametric estimates of the search cost distribution well. Observe that the estimation error term in the pricing equation (the estimated counterpart of (9)) can be interpreted as the unobserved marginal costs of the firms, leading to a pure strategy equilibrium as discussed in footnote 21.

[27] As discussed in detail in the Data Appendix, there is a conceptual issue regarding how market shares should be measured in this industry. Asset (stock-based) shares, our primary measure, seem in many ways reasonable and the necessary data is readily available. However, they differ from typical market share measures defined in terms of purchase flows. Unfortunately, we do not observe gross purchases, only net flows into a fund. We instead construct our flow-based quantity measure by summing all positive monthly net flows throughout the year. This approximate gross inflow measure yields market shares highly correlated (a coefficient of 0.93) with the asset-based shares.

load funds are invariant to the investor's holding period (since they charge only asset-based annual fees), annualized costs for load fund investors depend on how long assets are held. For example, annualized costs for a typical back-end load fund are highest for redemptions made in the purchase year, but decline with each year the assets are held. Since we do not have information regarding how long investors hold their assets in these funds, and how much of the load fees they actually pay, in columns (C) and (D) we ignore all loads and use only funds' annual management fees as our price measure, which is a rather optimistic lower bound on what investors actually pay. In the specification of column (E), we allow Z_j to be a linear combination of both the fund's age and the logged number of funds managed by the funds' management company. (The age coefficient is normalized to one and an estimated coefficient γ multiplies the logged number of funds.) The intent is to account for the possibility that some investors consider S&P 500 index funds while already holding assets in non-sector funds. If investors are more likely to purchase sector funds already in the same fund family as the funds which they already own (say because of switching costs as discussed above), including a measure of the size of the fund family in the search probability is a way to proxy for this effect.

Observe first that the parsimoniously parameterized model explains both price and market share extremely well. In our benchmark model, we explain 92 percent of price variation and 98 percent of the variation in market shares. The corresponding values for the other specifications are also high. Notice, too, that the estimates' qualitative features are consistent across the specifications. There is little quantitative difference in the estimates either, with the possible exception of the specification using flow-based market shares.

A few patterns regarding the search cost distribution are evident in the results. The estimated search costs appear reasonably sized. Furthermore, the mean logged search cost is trending downward throughout the sample period, while at the same time variance is increasing. We shall return to this result in detail below. The benchmark results imply a median search cost of 5 basis points (a $5 search cost per $10,000 of assets invested) in 2000, down from 21 basis points in 1996. There is considerable heterogeneity in search costs across the population, and the distribution is highly skewed; the twenty-fifth- and seventy-fifth-percentile search costs are 0.7 and 75 basis points, respectively. The flow-based market share specification implies lower search cost levels, but exhibits the same trends. The "lower-bound" (annual fees only) price measures used in (C) and (D) yield lower median search cost estimates than those from 7-year, load-inclusive annualized prices. This makes sense because there is less dispersion to explain in annual fees alone. Hence to the extent that the literature-standard price is an upwardly biased measure of the true fees paid by load fund investors, estimated search costs using the standard measure will be overstated. It does not appear that any mismeasurement of prices paid by load fund investors is greatly influencing the other results, however. The trends in the mean and variance of search costs remain and other parameter estimates change little.

Our estimates imply substantial asymmetries in fund-sampling probabilities are necessary to explain the data if funds are equivalent in all but price. The benchmark estimate of α is 2.6, indicating that fund age increases market exposure positively. This implication is supported in the data: there are 11 pairs and 8 triples of funds with equal prices and unequal ages in our sample, and the older (oldest) fund has the largest market share in 8 of the pairs and 6 of the triples. The estimated age effect is also quite nonlinear (for example, an investor has a 54 percent higher probability of finding a fund one year older than one of average age in 2000). This is to a great extent driven by the model wanting to match the dominant market share of Vanguard, the oldest fund. In model (E), where the sampling probability is a function of both fund age and the number of funds run by its management company, γ is positive and significant. Ceteris paribus, then, sector funds in larger fund families are more likely to be found (and purchased) by investors.[28] We return to this point below. The estimated marginal costs of fund companies are small; the benchmark estimate of the mean is 4 basis points (i.e. it costs \$4 to administer an additional \$10,000 in assets).

V.C. Heterogeneous Funds—Implied Search Costs

Letting sampling probabilities vary with fund attributes allows the model to fit the data quite successfully, and produces what are in our opinion several economically sensible results. However, an alternative explanation for very different market shares among similarly priced funds is that they might simply be different products. To incorporate non-price differentiation

[28] We have also estimated specifications (not reported here) where, besides age, the fund's distribution channel—whether it is a load or no-load fund—can also affect its sampling probability. Since load funds are sold through brokers rather than directly to investors, we thought it possible that the two types of funds might have systematically different sampling likelihoods. We found that the estimated load coefficient was economically small and its inclusion barely changed the estimate of α. As we show in the next section, the load/no-load distinction does seem to matter, but in a different way than through sampling probabilities.

into the model, we assume that sampling probabilities are equal.[29] As shown above, the search cost distribution as well as fund utilities u_j are nonparametrically identified in this case. We use the observed fund market shares and prices to trace out the implied search cost distributions using the procedure discussed in Section IV.2. We assume funds' marginal costs are identical and equal to 10 basis points. Experimenting with marginal costs as large as 30 and as low as 0 basis points yielded similar findings.[30]

We plot in Figure IV the nonparametrically identified search cost quantiles for 1996 and 2000 along with log-normal distributions fit by least squares to these quantiles.[31] While we can identify separate search costs distributions for each year in our sample, we only plot two years to ease clutter. The 1995 distribution is similar to 1996, and the 1997 through 1999 distributions indicate a more-or-less continuous evolution in the direction of the plotted 2000 distribution.

The results highlight a considerable decline in search costs for investors below the eighty-fifth percentile of the distribution. For example, the median search cost in 1996 (interpolated from the fit parametric curve) is 1.5 basis points, while the 2000 median is 0.2 basis points. Over the same period, however, search costs appear to have diverged across the population: the high end of the distribution saw increases in search costs. Notice that these results are consistent with those from the unequal-sampling-probabilities specification above (although the median search cost levels are smaller here). There, average logged search costs fell but their variance grew over the period. Calculating the search cost levels implied by that specification indicates that, as seen here in the differentiated-product model, search costs at the far end of the distribution actually increased despite the falling mean. (Although the previous results imply increases only above the ninety-seventh percentile, rather than the eighty-fifth, as is

[29] In reality, funds are likely to both be differentiated and have different sampling probabilities. However, as evident from our discussion of identification, it is not possible to identify both effects separately using data on prices and market shares only. We must therefore investigate each possibility in isolation.

[30] This assumption of identical marginal costs (which we make since we do not have data on individual funds' marginal costs) raises the issue of the existence of pure strategy equilibria with price dispersion. In this case, however, funds are heterogeneous in the utility levels they offer. This heterogeneity may well be akin to a "productivity heterogeneity" that would help sustain a dispersed-price equilibrium. We do not have a formal result proving this supposition, however. We instead verified (using our estimating the demand parameters) that funds were indeed setting their "best-response" prices given their competitors' prices and offered utility levels. We found that the estimated profit functions were single-peaked and attained their maxima near (within a reasonable statistical discrepancy) the observed prices. See Hortaçsu and Syverson [2003] for details. Other empirical contexts may prove more problematic, however, and we caution future researchers utilizing this estimation strategy not to skip this "best-response verification" step.

[31] The price and market share data utilized in this estimation is the same as in Table III, column (A). We found that data specifications corresponding to columns (B) through (D) yielded similar results.

the case here.)

Intriguingly, this divergence in estimated search costs across investors was concurrent with a documented influx of novice mutual fund investors. It seems likely that these new investors had significantly higher search costs than prior investors, and/or they would also be more likely to desire the services of a financial advisor (which come bundled with load funds) when buying funds. If so, a compositional shift in the investor population might explain the changes in the estimated search cost distributions as well as the within-sector market share reallocations described above. We investigate this possibility in more detail below.

V.D. Heterogeneous Funds—Fund Attributes Valued by Investors

We now estimate the contribution of funds' observable characteristics to investor utility using (15). We include ten attributes in X_j. There are dummies indicating whether a fund charges a load, and if it is a rear or deferred load. Loads are a pricing element (which we have already amortized into the price measure), but they also indicate funds sold with bundled broker services that investors may value. Rear or deferred loads indicate the presence of formal switching costs to removing assets from the fund. We also include a dummy if the fund is an exchange-traded fund (i.e., SPDRs or Barclay's iShares) to control for the special liquidity and intra-day pricing features of ETFs. We measure the number of additional share classes attached to the fund's portfolio; for a single-share-class fund this value is zero. The number of other funds managed by the same management company is included to capture any value from being associated with a large fund family. Fund age is in the regressions as well. (Here, both the number of family funds and age enter in logs to parsimoniously embody diminishing marginal effects. Recall that we instrument for age because of its possible correlation with unobservable quality.) We add the current fund manager's tenure, measured in years, as a covariate. And while all of the funds in our sample seek to match the return profile of the S&P 500 index, they do exhibit some small differences in their financial characteristics. These can result from skilled trading activities by a fund's management despite having a severely constrained portfolio. We thus include measures of tax exposure (the taxable distributions yield rate), the yearly average of the ratio of monthly fund returns to those of the S&P 500 index, and the standard deviation of monthly returns. To the extent that fund buyers prefer any persistent positive variations in financial performance, these controls should capture much of this effect.[32]

The utility function results are presented in Table IV. (We impose that the utility loadings β are constant across our sample years.) The results are qualitatively sensible. The coefficient on the exchange-traded fund dummy is positive and significant, as are the number of funds in the fund family, and the fund-S&P 500 index return ratio. Furthermore, higher tax exposure affects utility negatively and significantly. The positive coefficient on logged fund age, while positive, is not significant at the 5 percent level. The number of other share classes sharing a common portfolio and manager tenure also have positive coefficients, although these impacts cannot be statistically distinguished from zero. These characteristics all enter the estimated utility function in the expected direction. The most puzzling result is that the standard deviation of monthly returns coefficient is positive and significantly so. Ceteris paribus, investors should prefer a fund with less return volatility, not greater.[33]

The coefficients on the load indicators deserve further attention. The "any load" dummy is insignificant, suggesting that bundled advice is not a vertical component of utility; i.e., an attribute that all investors value. (We consider the possibility that there are horizontal preferences for such advice below.) The positive coefficient (although only significant at the 11 percent level) on the rear/deferred load dummy is unusual given that it indicates barriers to asset withdrawals. However, revisiting the above discussion on identification suggests an interpretation. Funds' indirect utility estimates are identified in part by market shares; the model explains the relatively large market shares of certain high priced funds by attributing to them a

[32] We also collected in a phone survey measures of funds' account service quality. These measures include the number of account statements per year, the fraction of the day account access by phone is available, a dummy indicating whether investors are able to write checks out of their fund balances, and a dummy indicating whether the fund provides phone access to financial advisers. We choose not to include these variables in our reported specifications because we were unfortunately unable to obtain historical account service data. However, when we estimated a utility function that included these measures a slightly smaller sample, we found that the number of statements and the availability of phone advisers had no significant measured utility effect, and that more account access by phone entered positively and significantly. The coefficient on check-writing privileges was negative and significant, strangely. Including these service measures did not greatly change the qualitative or quantitative utility impacts of the other attributes discussed below.

[33] It has been suggested to us that this result may reflect unmeasured liberalness of funds' "in-and-out" privileges. If a fund catered to day traders, say, the fund would likely have both a high standard deviation of monthly returns (because meeting redemption demands would make it more difficult for the manager to track the S&P 500 index) and a higher market share than predicted by the model. We attempt to indirectly test for this mechanism by regressing funds' standard deviation of monthly returns on a set of year dummies and the absolute value of funds' monthly asset flow rates (i.e., net flows into the fund during the month divided by average of fund assets at the beginning and end of the month) averaged over the year. We do not find the positive correlation between return variability and flow rates implied by in-and-out-privileges story. The estimated flow rate coefficient is significantly negative at the 1 percent (10 percent) level when fund fixed effects are excluded from (included in) the regression.

high utility level. The positive rear/deferred dummy suggests that such funds have larger market shares than is implied by their other attributes. This is probably because these funds have built-in switching costs. Some assets remain in these funds not because they have other favorable attributes, but instead because it would be costly for investors to remove them. As such, this estimate is evidence of switching costs effects in the sector, at least for certain funds.

V.E. A Possible Explanation for the Changes in Search Costs

As we suggest above, the shifts seen over the sample in the search cost distributions may result from a composition shift in sector investors. How might this happen? An influx of novice investors, less financially savvy than those already in the market, enter the sector during the sample period. Their lack of experience has two implications for their search behavior vis-à-vis more experienced investors: their information costs are higher on average, and they are more likely to value the financial advisor services that come with load funds. This leads to increases in search costs at the upper quantiles of the search cost distribution even as most investors enjoy reductions in search costs because of technological improvements or experience effects. Further, it implies the increases at the upper end of the distribution should be especially stark for load funds, since they are more likely to be purchased by the high-search-cost neophytes. It also entails growth in the total market share of load funds, of course.

There is suggestive evidence for this mechanism. The asset growth in the retail S&P 500 index fund sector did occur at a time when a large number of households participated in the mutual fund market for the first time. According to the Investment Company Institute [various issues, 1996-2000], participation rose from 37.2 percent of households holding at least one fund in 1996 to 49 percent by 2000, bringing 14.9 million new households into the mutual fund market.

These new investors were different than those already invested in mutual funds. An ICI [2001] profile of mutual fund shareholders shows that new investors (those having purchased their first mutual fund since 1998) are significantly younger; less educated; and have lower incomes, financial assets, and experience levels with other investment products than "seasoned" investors (those buying their first fund before 1990).[34] Over a third of new investors assessed their understanding of mutual fund investing to be "limited or none." Accordingly, a higher

[34] Similar differences were observed in other surveys conducted by ICI in earlier years.

fraction of these new investors stated a reliance on the advice of professional financial advisers when making fund purchases and sales decisions than did seasoned investors. They were also only half as likely as seasoned investors to have reported purchasing a no-load fund (either directly from a fund company or through a discount broker). This survey evidence is consistent with the above characterization of novice investors as high-information-cost investors who are more likely to buy load funds in order to receive the financial advice bundled with them.

We try to (at least partially) capture the effects of this mechanism by estimating a specification of our model where investors differ by type in terms of their preferred distribution channel. That is, one group seeks to purchase load funds through an advice-providing agent, and the remainder buy no-load funds directly.[35] (Thus the fraction of each type in the investor population is simply the share of assets held in each type of fund.) Note that this specification, unlike those above, incorporates a form of horizontal differentiation. Although simple, it is of particular interest given the discussion above and the institutional setup of the industry. We allow full flexibility across the two groups' search cost distributions. Estimation proceeds as before, except search costs are estimated for investors preferring each purchase channel separately, using only those funds in the respective channel.

Figure V plots the resulting estimated search cost distributions for both investor groups in 1996 and 2000. (Again, while the exercise can be repeated for each of the years in our sample, we only plot the 1996 and 2000 distributions for the sake of comparison to the results in Section V.C, and because these years span the period the asset reallocation to new, high-priced funds occurred.) As can be seen, the search costs of no-load fund investors are much lower than their counterparts who buy load funds. This is not surprising; investors who buy load funds through the "sales force" are likely to have higher information-gathering costs in financial matters.[36] Furthermore, it is also apparent that while search costs decreased over the sample for nearly all no-load fund investors, they increased for the top 40 percent of load fund investors. This repeats the pattern seen above: search cost decreases at the low end of the distribution with increases at the high end. Here, though, the decomposition shows it is almost exclusively manifested within

---

[35] While novice investors are more likely to be load fund investors, a substantial fraction of seasoned investors report owning load funds as well. Hence one should not interpret the group of load fund buyers as exactly equivalent to the set of novice investors.

[36] The ICI's [2001] investor profile also provides supporting evidence for this. Load fund investors are less educated, have lower incomes, and less likely to have Internet access than no-load investors. Of course, some of these differences are driven by the different investing experience profiles of the respective groups.

load fund investors. This is as suggested above—a number of high-information-cost inexperienced investors entered the market and bought loaded S&P 500 index funds. Further bolstering this notion is the fact that most of the entry and market share gains at the high end of the price distribution (documented earlier) was accounted for by load funds (indeed, their market share within the sector nearly doubled from 1996-2000).

We emphasize that our model's implication of such a composition shift is only suggestive—we would need investor-level data to test it definitively—but we do find the story tantalizing as well as consistent with much of the evidence on events in the sector and the characteristics of its likely investors.

VI. Welfare Implications

Since most investors' portfolio needs could be met by a single S&P 500 index fund, it is conceivable that having scores of sector funds harms social welfare by inducing wasteful search by investors and losses of scale economies in managing assets. We briefly consider here what our estimates imply about social welfare in the observed equilibrium. To do so, we compute the implied welfare change (arising from several sources) that is induced when a counterfactual market structure is imposed on the sector. These calculations, detailed in the Computational Appendix, are based on estimated from the model above as well as on estimates obtained using supplemental data and models. We wish to make it clear from the outset that our model, as is typically the case, is stylized and therefore omits some institutional details of the market that may impact welfare. Hence these calculations are only suggestive.

To make the exercise transparent as possible, we consider a polar-case counterfactual: entry into the sector is restricted to only the Vanguard 500 Index Fund, the first and still dominant player in the sector. While this counterfactual is perhaps not particularly realistic, its stark nature underscores the nature and size of the possible welfare impacts the many sector funds have had.[37]

[37] Economists have long recognized that free entry can be socially inefficient. Mankiw and Whinston [1986] formalize this notion in a model where entry creates a loss of scale economies due to the over-spreading of output across production units. Stiglitz [1987] and Stahl [1989], in theoretical exercises of particular relevance to our framework, highlight socially harmful effects on buyer search behavior as the number of market producers increases. Of course, there are counterbalancing positive influences of entry on social welfare. Entry can reduce distortions of market power by increasing incumbents' demand elasticities, depending on the particulars of the search process. Product variety is the second possible benefit of entry, if goods are heterogeneous and consumers differ in how they value product attributes. We consider these in turn.

There are four major welfare manifestations of the large number of sector funds. One is through search. An increase in the number of funds in our search equilibrium can create welfare losses in two ways: the loss that occurs when investors end up purchasing funds that do not offer the highest indirect utility, and investors' direct expenditures on learning about funds' attributes. A separate welfare impact is through the possible loss of scale economies that occurs when output is overspread across production units. The third welfare effect—one that increases with the number of funds—comes from product variety benefits. While our model implies no welfare loss above from restricting entry to Vanguard alone (because it already offers the highest indirect utility among all funds), this result is an artifact of assuming purely vertical differentiation and no outside good. If horizontal differentiation exists, restricting the number of sector funds may well lead to product-variety welfare losses. We estimate a standard logit demand model [e.g., McFadden 1974], which incorporates horizontal taste differences, to estimate the size of these losses. The final welfare impact, of course, is the effect that entry has on competition and the reduction of market power. Granting a monopoly (an unregulated one at least) could well create greater distortions from market power.

Table V summarizes the welfare calculations. It shows in basis points and dollar values the total welfare changes that would be induced by the imposition of a Vanguard monopoly. The welfare gains from eliminating search are shown in columns (3) and (6). As can be seen, while the proportional gains (i.e., per dollar of assets) fell during the sample, asset growth in the sector led to increases in absolute gains until 2000. The cost savings from preserving scale economies in fund operations grew throughout the sample, as this is directly related to the number of funds—see column (5). At the same time, however, the monopoly counterfactual's product variety welfare *loss* also climbed. As seen in column (8), these results imply that on net there are growing welfare benefits of restricting entry throughout 1995 to 1999 with a small loss implied in 2000.

To see these values from the perspective of the prices investors face, as well as to give an idea of the potential competitive effect of entry necessary to negate the welfare gains in the counterfactual, we calculate in column (9) the increase in Vanguard's price that would leave the average investor indifferent between the observed market structure and the imposed monopoly. (We keep things simple here by assuming inelastic demand, so no assets would leave the sector in response to a price change.) Considering that Vanguard's price is at or below 20 basis points during our sample period, the price changes required to make investors indifferent are substantial. There may be nontrivial welfare benefits of imposing a monopoly.

These calculations come with a host of caveats, however. We do not directly include welfare losses that may arise due to the deadweight loss of monopoly, or any transfer in surplus from consumers to producers if there are distributional concerns. While such market power losses could plausibly be avoided by regulating the monopolist, regulation induces its own well known welfare costs (rent seeking, corruption, etc.) that might be themselves substantial. Also, the product variety welfare and fixed cost savings estimates were made using assumed functional forms that, while facilitating straightforward interpretation (see the appendix), yield estimates that should be considered bounds rather than point estimates. The true product variety welfare loss in the monopoly counterfactual may be less than estimated, while fixed operating cost savings may be overstated. These tend to cancel each other out, so their net effect depends on the departures of the actual welfare impacts from these calculated bounds.[38]

Given these cautionary notes, we are reluctant to prescribe policy based on our welfare findings. They do not strike us as unreasonable, however. As such they may offer guidance in thinking about the welfare impacts of search and product differentiation in the retail S&P 500 index fund sector in particular and the mutual fund industry as a whole.

VII. Conclusions

We have presented evidence that key features of the U.S. mutual fund industry are driven by factors beyond financial portfolio heterogeneity alone. We focus on an asset segment, retail S&P 500 index funds, where all funds are characterized by nearly homogeneous return patterns. Despite the homogeneity, we find that this sector exhibits the fund proliferation and price dispersion patterns seen in the broader industry.

[38] We also thank a referee for pointing out that with horizontal differentiation, there would simply be less money in the S&P 500 index fund sector if only Vanguard's fund was available. The logit demand system includes an outside good, which we define as the aggregate of all other retail growth and income funds. Thus our welfare calculation for product variety effects accounts for this effect. For the search and fixed cost components, we can construct a lower bound for the counterfactual total market size by assuming that the Vanguard monopoly retains only its own assets, without capturing assets from its competitors (i.e., assuming away any demand spillovers). We recalculated Table V using this "lower bound" total market size under the counterfactual, and we found that the sign of the net welfare and price changes remained the same, although the positive welfare changes through 1995-1999 were smaller, and the negative welfare change in 2000 was larger. For example, the lower-bound figure for the "indifference price change" in 2000 is -12 instead of -1.2 basis points. For 1995, the price change was 33 basis points rather than 66 basis points.

We consider a combination of non-financial fund differentiation and information/search frictions as explanations for these observations. We construct an equilibrium model where investors with heterogeneous search costs shop over differentiated funds, and these funds compete with each other in prices mindful of investors' search behavior. The model is estimated using data from funds in the retail S&P 500 index fund sector, and the search costs necessary to sustain the observed price dispersion are found to be of reasonable magnitude. Indeed, the estimated search costs exhibit much less dispersion than the price variation they support. Standard models involving search alone, however, are rejected by the data. Differentiated funds, either in their marketplace exposure or in the vertical component of utility that they offer investors, are necessary to our results. We furthermore decompose the contribution of various fund attributes to investors' indirect utilities, and find qualitatively sensible utility weights.

Our estimated search cost distributions also shed light on sector developments over the course of our sample. We observe considerable entry of high-price funds into the sector accompanied by a concurrent shift in assets toward more expensive funds. This, despite the fact that technological improvements over the same period were probably lowering average information-gathering costs. The estimated search cost distributions offer an explanation for these seemingly divergent features. We find that while average search costs were declining, costs for those at the upper percentiles of the distribution actually tended to increase through our sample years. This widening of the distribution was concurrent with the documented increase in households' first-time participation in the mutual fund market, suggesting novice investors with high information/search costs caused this shift of assets into higher-price funds and supported the high-fee entrants. This is further supported when we allow investors to be horizontally differentiated by their preferences over funds' distribution channels; i.e., no-load (direct sales) versus loaded (broker-sold) funds. While virtually all investors considering no-load funds exclusively enjoyed decreases in search costs over the sample, the top 40 percent of the load fund investor distribution experienced search cost increases over the same period. This is consistent with the novice-investor composition shift occurring mainly in the load fund sector, a sensible implication if new investors place a higher value on the financial advisory services that come bundled with load funds.

We also consider the welfare implications that search costs and differentiated products pose when there are a large number of financially identical funds. We find that the total costs of the search process, both direct and indirect (through investors purchasing lower-utility funds), are sizeable. While these costs could be avoided by restricting entry into the sector to a monopolist fund, this policy could well cause its own welfare losses due to the reduction in product variety and increased market power. Our rough calculations indicate that imposing monopoly might be socially beneficial on net. However, given the myriad simplifying assumptions that are required in such calculations, we are reluctant to make any policy recommendations too strongly.

While we focus here on a particular mutual fund asset class to control for financial performance heterogeneity while highlighting the possible roles of search and non-portfolio product differentiation, we think our results also offer at least partial explanations for the fund proliferation and large fee dispersion seen in the mutual fund industry as a whole. They also suggest that large increases in household participation in mutual fund markets can have important and interesting impacts on the industry's market structure. Much more work needs to be done, however, to fully characterize these effects.

Computational Appendix

A. Market Share Equations

The market share of a particular fund depends on the probabilities that investors with search costs less than or equal to its corresponding cutoff search cost value from (4) finds the fund before finding another fund with a critical value higher than their search costs.

More formally, consider an investor with search cost $\hat{c}$. Then the probability that our investor ends up purchasing a particular fund k is equal to either (a) zero if $c_k > \hat{c}$, since the investor will always keep searching upon drawing fund k, or (b) if $c_k < \hat{c}$, the probability that he draws k before drawing any other fund with a critical value less than $\hat{c}$. The latter probability is equal to ρ_k times the sum of the probabilities of draw sequences where all funds have $c_k < \hat{c}$. We can express these probabilities, and therefore the corresponding market share equations (7), compactly using combinatorics. We derive these equations below.

First, define $\hat{c}_l$ as the largest fund cutoff search cost value that is less than $\hat{c}$:

$$\text{(A.1)} \qquad \hat{c}_l \equiv \max\{c_j \mid c_j < \hat{c}\},$$

An investor with search cost $\hat{c}$ will stop searching once a fund with $c_j \leq \hat{c}_l$ is drawn, because at this point the benefit of additional search is less than its cost. Consider drawing such a fund as a "success" event. Then the probability of T failures (i.e., where all T draws are of funds with $c_j > \hat{c}_l$) is equal to

$$\text{(A.2)} \qquad \sum_{\bar{a}|a_1+\cdots+a_{ll}=T}\left[\frac{T!}{a_1!\ldots a_l!}\rho_1^{a_1}\ldots\rho_l^{a_l}\right],$$

where the l subscript denotes the fund with $c_j = \hat{c}_l$, ρ_j is the probability of sampling fund j, a_j is the number of times that the fund is drawn in the sequence of T draws, and the sum is taken over all combinations of the a vector $[a_1,\ldots,a_l]$ that sum to T.

It is known from combinatorics theory that the above summation simplifies to $(\rho_1 +\ldots+ \rho_l)^T$. Therefore the probability that fund k is chosen after T failures—that k is the "success" draw—is $\rho_k(\rho_1 +\ldots+ \rho_l)^T$. This expression must be summed over values of T (from $T = 0$, an immediate success where k is the first fund drawn, to $T = \infty$, the limit of possibility) to obtain the total probability that an investor with search cost $\hat{c}$ purchases fund k (alternatively, the probability that k is chosen given $c_k \leq \hat{c}_l$). That is,

$$\text{(A.3)} \qquad \Pr(\text{Fund } k \text{ chosen} \mid c_k \leq \hat{c}_l) = \rho_k \sum_{T=0}^{\infty}(\rho_1 + \ldots + \rho_l)^T = \frac{\rho_k}{1-\rho_1-\ldots-\rho_l}.$$

This is the probability (b) in the second paragraph above.

Of course, this probability depends on the value of $\hat{c}$, since this determines the particular value of ρ_l. Investors with differing search costs thus have various probabilities of finding a "success" in fund k. For example, the highest-search-cost investors (i.e., those with $\hat{c} > c_1$) have $\hat{c}_l = c_1$, so they only take one fund draw and purchase whichever fund they find. Thus the contribution to fund k's market share from investors with search costs above c_1 is $\rho_k[1 - G(c_1)]$. Analogously, for all $k \geq 2$, the market share comprised of investors in the next-lower segment of the search cost distribution (between c_2 and c_1) are those investors with that only draw fund 1 until fund k is chosen. This market share is then $[G(c_1) - G(c_2)]\rho_k/(1 - \rho_1)$. These are the values discussed in the text for $k = 2$.

This is easily generalized. Investors with search costs between c_3 and c_2 purchase k if $k \geq 3$ and if k is the first fund drawn besides fund 1 or 2. Thus their contribution to the market share of fund k is $[G(c_2) - G(c_3)]\rho_k/(1 - \rho_1 - \rho_2)$. Generically, investors with search costs between cutoff values c_j and c_{j-1} account for a market share for fund k equal to the following, if $k \geq j$:

$$\text{(A.4)} \qquad \frac{\rho_k}{1-\rho_1-\ldots-\rho_{j-1}}\left[G(c_{j-1})-G(c_j)\right].$$

The total market share for k is then the sum of these values for all segments of the search cost distribution above c_k:

$$\text{(A.5)} \qquad q_k = \rho_k\left[1-G(c_1)\right]+\frac{\rho_k}{1-\rho_1}\left[G(c_1)-G(c_2)\right]+\ldots+\frac{\rho_k}{1-\rho_1-\ldots-\rho_{k-1}}\left[G(c_{k-1})-G(c_k)\right].$$

Grouping common $G(c)$ terms, factoring out a ρ_k, and evaluating at $k = j$ yields expression (7) in the text.

<u>B. Derivatives of Demand Curves</u>

The generalized market share equations are given in (7):

$$\text{(A.6)} \qquad q_j = \rho_j\left[\begin{array}{l} 1+\dfrac{\rho_1 G(c_1)}{1-\rho_1}+\dfrac{\rho_2 G(c_2)}{(1-\rho_1)(1-\rho_1-\rho_2)}+\displaystyle\sum_{k=3}^{j-1}\dfrac{\rho_k G(c_k)}{(1-\rho_1-\ldots-\rho_{k-1})(1-\rho_1-\ldots-\rho_k)} \\ -\dfrac{G(c_j)}{(1-\rho_1-\ldots-\rho_{j-1})} \end{array}\right].$$

We want to take price derivatives of these equations. Notice is where prices enter into the equations: the indirect utilities provided by the funds include their prices, and these indirect utilities are in turn embodied in the cutoff search cost values $c_1,\ldots,c_j$ above.

So with this in mind we can take the price derivative of the above:

$$\text{(A.7)} \qquad \begin{array}{l} \dfrac{dq_j}{dp_j}=\dfrac{\rho_1\rho_j g(c_1)}{1-\rho_1}\dfrac{dc_1}{dp_j}+\dfrac{\rho_2\rho_j g(c_2)}{(1-\rho_1)(1-\rho_1-\rho_2)}\dfrac{dc_2}{dp_j}+\displaystyle\sum_{k=3}^{j-1}\dfrac{\rho_k\rho_j g(c_k)}{(1-\rho_1-\ldots-\rho_{k-1})(1-\rho_1-\ldots-\rho_k)}\dfrac{dc_k}{dp_j} \\ \qquad -\dfrac{\rho_j g(c_j)}{(1-\rho_1-\ldots-\rho_{j-1})}\dfrac{dc_j}{dp_j} \end{array}.$$

Now recall from (4) that

$$\text{(A.8)} \qquad c_j = \sum_{k=j+1}^{N}\rho_k\left(u_k-u_j\right).$$

Combined with the fact that the derivative of the u with respect to price is –1, this implies

$$\text{(A.9)} \qquad \frac{dc_r}{dp_j}=\begin{cases} -\rho_j, & \text{if } r<j \\ \displaystyle\sum_{k=j+1}^{N}\rho_k, & \text{if } r=j \\ 0, & \text{if } r>j \end{cases}.$$

Substituting this into (A.7) gives

(A.10) $$\frac{dq_j}{dp_j} = -\frac{\rho_1 \rho_j^2 g(c_1)}{1-\rho_1} - \frac{\rho_2 \rho_j^2 g(c_2)}{(1-\rho_1)(1-\rho_1-\rho_2)} - \sum_{k=3}^{j-1} \frac{\rho_k \rho_j^2 g(c_k)}{(1-\rho_1-\ldots-\rho_{k-1})(1-\rho_1-\ldots-\rho_k)} - \frac{\rho_j g(c_j) \sum_{k=j+1}^{N} \rho_k}{(1-\rho_1-\ldots-\rho_{j-1})}.$$

This is equation (10) in the text.

In the special case of equal sampling probabilities ($\rho_j = 1/N \; \forall \; j$), this simplifies to:

(A.11) $$\frac{dq_j}{dp_j} = -\sum_{k=1}^{j-1} \frac{g(c_k)}{N(N-k+1)(N-k)} - \frac{(N-j)g(c_j)}{N(N-j+1)}.$$

C. Welfare Calculations

The model implies two potential welfare losses from search when there are a large number of funds. One is from investors ending up at a fund with lower utility ranking than *N*—i.e., the (negative of the) utility gain that would occur if the fund with the highest utility value (Fund *N*) had 100 percent market share. This is easily calculated as the market-share-weighted utility difference:

(A.12) $$Loss_1 = \sum_{j=1}^{N} q_j (u_N - u_j) = u_N - \sum_{j=1}^{N} q_j u_j \;,$$

where the utility levels used in the calculation are those measured in the nonparametric specification above. Using the observed market shares and the estimated indirect utility levels from Section V.3., we find the calculated losses fell from 42.1 basis points in 1995 to 15.6 basis points in 2000. These costs are not huge—not surprising since the highest-utility fund holds a majority position—but at the same time are not trivial. Furthermore, given that the absolute level of losses is equal to the product of these proportional losses and total sector assets, the estimated losses in dollar terms run into the several-hundred-millions.

The second search-related loss, that from direct search cost expenditures, is simply equal to total search expenditures in the observed free-entry equilibrium, since no search would be necessary with entry restricted to a single fund as in the counterfactual. Using a similar line of argument as in the derivation of (A.2) and (A.3), the expected number of searches conducted by investors with search costs $c_{k+1} < c < c_k$ can be calculated as the mean duration of a geometric failure model. This number, multiplied by the incurred search cost, is then integrated over the population of investors to yield the resulting consumer surplus due to search. In practice, we evaluated this integral numerically, taking 100,000 draws ("investors") from the estimated search cost distribution. We found that the average number of funds searched by investors grew in conjunction with the number of sector funds, increasing over 2.5 times (from 9.3 to 24.8) from 1995 to 2000. However, total search expenditures dropped by nearly two-thirds (from 23 to 8 basis points) because average search costs were falling throughout the period.

While the vertical differentiation assumption of the model implies no product variety gain from having a large number of funds other than Vanguard, we can approximate the possible product variety benefits if horizontal differentiation is important by assuming that sector demand is characterized by the standard logit demand system. To model the utility benefit of not owning an S&P 500 fund, we have to define an "outside good" and compute market shares using this definition. We defined total market size in a year as total assets in all growth and income funds in CRSP, prorated by the share of retail S&P 500 funds' assets in the S&P 500 funds total. We then estimate the logit demand system following Berry [1994]. Besides having the benefit of being computationally simple to implement, the logit model has well-known biases that tend to make it overestimate variety benefits. (Petrin [2002] contains a discussion of welfare measurement biases in standard logit models.) Thus we can interpret the welfare effect implied by the logit as an upper bound on the variety loss of the monopoly counterfactual.

We follow Small and Rosen [1981] in constructing the welfare loss from restricting the choice set to Vanguard. We find attribute utility weights implied the logit estimation to be qualitatively similar to those obtained in the utility-weighting estimates in Section V.3 above. The estimated coefficients have the same signs in both models in all but one case, which happens to be the most puzzling of the prior specification (the monthly return standard deviation). This attribute coefficient has the expected negative sign here. The implied product variety welfare benefits are discussed in the text.

To estimate the potential social losses from the loss in scale economies due to over-spreading output across the many funds, we estimate a simple annual cost function. The function is linear in funds' total net assets; i.e.,

(A.13) $C_t = F + mc \cdot TNA_t$.

That is, the total cost of operating a fund during a given year (i.e., "production" costs) are equal to a fixed cost F plus the product of the fund's total net assets under management and a constant marginal cost mc. We assume both F and mc are constant over the time period of our sample. We estimate (A.12) using total cost data that we obtained from Lipper Analytical Services for a subsample of our funds (approximately 80 percent of the fund-year observations). This information is gathered from funds' annual expense statements.

While extremely simple, this specification offers an easily characterized welfare loss of entry through the loss of scale economies. The social cost of an additional fund is simply the fixed cost, because any assets can be shifted among funds at a the same marginal cost. If returns to scale in the industry were alternatively characterized by declining marginal costs, for example, this loss is much harder to measure, because it would then depend on how many assets are taken from which funds and those funds' initial assets levels. Despite its over-simplicity, we believe it is a reasonable first-order approximation.

The cost function estimates imply an annual fixed cost of \$1.22 million (s.e. = \$0.16 million) and a marginal cost of \$0.00182 per dollar of assets—18.2 basis points (s.e. = 0.1 basis points). The coefficients are estimated fairly precisely, particularly for marginal cost, and the goodness of fit is high (R^2 = 0.98). Both coefficient estimates strike us as reasonable. The resulting welfare gain from the loss of scale economies in our counterfactual is simply the sum of the fixed costs of the extra funds. In reality, it is likely that a single index fund with such a large amount of assets could run up against increasing marginal costs, suggesting then that these estimates should be considered an upper bound.

Data Appendix

The bulk of our performance and characteristic data on mutual funds comes from the CRSP mutual fund database for the years 1995-2000. This data includes a considerable amount of information about the funds (as mentioned previously, different share classes for the same asset pool are considered separate funds). Most of the data is compiled annually, but monthly information on returns and assets under management is also available. We observe the year the fund was established, the identity of the fund's manager and managing company, the starting date of the manager's tenure, and whether the fund enters or exits a given year. Annual performance and portfolio characteristics included in the data are income and capital gains distributions. Pricing information includes expense ratios; 12b-1 fees (like expense ratios, these are annual fees, but they are specifically earmarked to cover fund marketing and distribution costs); and front, rear, and deferred load levels. We use this data to compute a number of supplementary variables for our analysis. These include fund age, the total number of funds managed by a fund's management company (including those outside the S&P 500 index category), the average and standard deviation of monthly returns, and measures of gross return benchmarked to the return of the S&P 500 index.

We supplement the CRSP data with mutual fund cost data from Lipper Analytical Services for 1995-2000. Lipper examines funds' year-end reports to gather cost data in a number of categories. These cost numbers are aggregated and combined with asset data to yield total annual costs. We have cost data for approximately 80 percent of our sample fund observations over 1995-2000.

Two measures are central to our empirical work: price and market share. Unfortunately, for mutual funds, neither concept is as straightforward as it often is for other products. Several central issues arise; we discuss those relevant to each measure in turn.

There are a number of ways in which mutual fund prices vary from traditional concepts of price. One of these is that, invariably, rather than being priced as a simple dollar level, mutual fund prices are fractional charges related to the size of asset flows into or stocks held in the fund. Furthermore, there are several dimensions along which mutual funds can be priced. All funds have an annual expense ratio. This is a percentage of investor assets that is withdrawn from the owner's account and used to reimburse the fund management company. Annual expenses include management, administrative, and in some cases 12b-1 (marketing and distribution) fees. Some funds also impose one-time loads, charged as a percentage of fund flows into or out of a fund. There are two types of loads typically employed by the industry. Front-end loads are charged at the time of a purchase of fund assets. Back-end (or deferred) loads, if applicable, are charged at a time of withdrawal.

Funds differ in both their chosen pricing instruments and their levels. Furthermore, because of the stock/flow distinction between annual fees and loads, as well as the timing discrepancy between front- and back-end loads, it is not a trivial matter to identify a single price for each fund. We use the approach, common in the literature, of measuring fund price by adding annual fees (the expense ratio) to one-seventh of the sum of all load levels. The one-seventh fraction is obtained from the stylized fact that a typical mutual fund account is held for about seven years. Loads are incurred only when there are flows into or out of a fund, not on any asset stocks held. Hence the price is meant to incorporate the shareholder's annualized cost of the load.

Market size is also less than empirically clear-cut in this industry. While market size measurement issues are not uncommon, they typically revolve around defining the boundaries of the market. In our case, however, this is secondary. What is more difficult is defining the unit of purchase. Assets may be a reasonable dimension along which to measure market shares, and the data is readily available. Using this measure implicitly assumes investors annually evaluate whether to continue holding their assets in a particular fund or to move them elsewhere. (This is further complicated by the fact that loads apply to flows but management fees apply to stocks.) Further, market shares in standard markets are typically defined as the share of total purchases accounted for by a producer; i.e., the flow, not the stock, as is the case with the asset-based measure. Unfortunately, our data does not allow straightforward measurement of gross fund purchases. Because we only measure total assets under management, we only observe net flows into a fund. Net flows hide the size of the gross flows underlying them, and gross inflows are the preferable measure to base flow-centered market shares upon. (For example, in our 2000 data, it was not uncommon to see negative net changes in fund assets over the course of the year. It is not exactly clear just how one could define a flow-based market share measure from these negative values.) This problem could be partially circumvented because we observe monthly asset and return data. Monthly net flows are then computed and summed by sign to obtain an aggregated measure of gross flows. We have run some specifications using these flow data (some of which are discussed in the paper) and found little qualitative difference from the asset-based measures in the results. We should also note that this approximate gross inflow measure yields market shares that are highly correlated with the asset-based measure (the correlation coefficient is 0.93).

University of Chicago and National Bureau of Economic Research
University of Chicago and National Bureau of Economic Research

References

Anderson, Simon P., André de Palma, and Jacques-François Thisse, *Discrete Choice Theory of Product Differentiation*, (Cambridge, MA: MIT Press, 1992).

Anderson, Simon P., Regis Renault, "Pricing, Product Diversity and Search Costs: A Bertrand-Chamberlin-Diamond Model." *RAND Journal of Economics*, XXX (1999), 719-735.

Barber, Brad M., Terrance Odean, and Lu Zheng, "Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows." Working Paper, Graduate School of Management, University of California, Davis, 2002.

Berry, Steven, "Estimating Discrete Choice Models of Product Differentiation." *RAND Journal of Economics*, XXV (1994), 242-262.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium." *Econometrica*, LXIII (1995), 841-890.

Burdett, Kenneth and Kenneth L. Judd, "Equilibrium Price Dispersion." *Econometrica*, LI (1983), 955-969.

Capon, Noel, Gavan J. Fitzsimons, and Russ Alan Prince, "An Individual Level Analysis of the Mutual Fund Investment Decision." *Journal of Financial Services Research*, X (1996), 59-82.

Carhart, Mark M., "On Persistence in Mutual Fund Performance." *Journal of Finance*, LII (1997), 57-82.

Carlson, John A. and R. Preston McAfee, "Discrete Equilibrium Price Dispersion." *Journal of Political Economy*, XCI (1983), 480-493.

Chevalier, Judith and Glenn Ellison, "Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy*, CV (1997), 1167-1200.

______, "Career Concerns of Mutual Fund Managers." *Quarterly Journal of Economics*, CXIV (1999), 389-432.

Eckstein, Zvi and Kenneth I. Wolpin, "Estimating a Market Equilibrium Search Model from Panel Data on Individuals." *Econometrica*, LVIII (1990), 783-808.

Falkenstein, Eric G., "Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings." *Journal of Finance*, LI (1996), 111-135.

Flinn, Christopher J. and James J. Heckman, "New Methods for Analyzing Structural Models of Labor Force Dynamics." *Journal of Econometrics*, XVIII (1982), 115-168.

Foster, Lucia, John Haltiwanger, and C. J. Krizan, "The Link Between Aggregate and Micro Productivity Growth: Evidence from Retail Trade." National Bureau of Economic Research Working Paper No. 9120, 2002.

Gabaix, Xavier and David Laibson, "Some Industrial Organization with Boundedly Rational Consumers." Working Paper, Harvard University, 2003.

Gruber, Martin J., "Another Puzzle: The Growth in Actively Managed Mutual Funds." *Journal of Finance*, LI (1996), 783-810.

Hong, Han and Matthew Shum, "Estimating Equilibrium Search Models." Working Paper, Princeton University, 2001.

Hortaçsu, Ali and Chad Syverson, "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds." National Bureau of Economic Research Working Paper No. 9728, 2003.

Investment Company Institute, *Fundamentals*, (Washington, DC: Investment Company Institute, Various Issues, 1996-2000).

______, *Understanding Shareholders' Use of Information and Advisers*, (Washington, DC: Investment Company Institute, 1997).

______, *Mutual Fund Fact Book*, (Washington, DC: Investment Company Institute, 2000 and 2002).

______, *Profile of Mutual Fund Shareholders*, (Washington, DC: Investment Company Institute, 2001).

Jensen, Michael C., "The Performance of Mutual Funds in the Period 1945-1964." *Journal of Finance*, XXIII (1968), 389-416.

Khorana, Ajay and Henri Servaes, "The Determinants of Mutual Fund Starts." *Review of Financial Studies*, XII (1999), 1043-1074.

______, "An Examination of Competition in the Mutual Fund Industry." Working Paper, Georgia Institute of Technology, 2003.

Malkiel, Burton G, "Returns from Investing in Equity Mutual Funds 1971 to 1991." *Journal of Finance*, L (1995), 549-572.

Mamaysky, Harry and Matthew Spiegel, "A Theory of Mutual Funds: Optimal Fund Objectives and Industry Organization." Yale International Center for Finance Working Paper No. 00-50, August 2001.

Mankiw, N. Gregory and Michael D. Whinston, "Free Entry and Social Inefficiency." *Rand Journal of Economics*, XVII (1986), 48-58.

Massa, Massimo, "Why So Many Mutual Funds?" Working Paper, INSEAD, 2000.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior." in P. Zarembka, ed., *Frontiers in Econometrics*, (New York: Academic Press, 1974).

Petrin, Amil, "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, CX (2002), 705-729.

Sirri, Erik R. and Peter Tufano, "Costly Search and Mutual Fund Flows." *Journal of Finance*, LIII (1998), 1589-1622.

Small, Kenneth A. and Harvey S. Rosen, "Applied Welfare Economics with Discrete Choice Models." *Econometrica*, XLIX (1981), 105-130.

Sorensen, Alan T., "An Empirical Model of Heterogeneous Consumer Search for Retail Prescription Drugs." Working Paper, University of California, San Diego, 2001.

Stahl, Dale O., "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review*, LXXIX (1989), 700-712.

Stiglitz, Joseph E., "Competition and the Number of Firms in a Market: Are Duopolies More Competitive than Atomistic Markets?" *Journal of Political Economy*, XCV(1987), 1041-1061.

van den Berg, Gerard J. and Geert Ridder, "An Empirical Equilibrium Search Model of the Labor Market." *Econometrica*, LXVI (1998), 1183-1221.

Table I
Price Dispersion within Fund Sectors

Sector	N	Mean Price	Coefficient of Variation	75th to 25th Percentile Ratio	90th to 10th Percentile Ratio
Aggressive growth	1278	191.0	0.485	2.0	3.1
Balanced growth	472	164.2	0.439	2.2	3.7
High-quality bonds	862	118.1	0.566	2.5	4.9
High-yield bonds	337	167.3	0.387	2.1	3.2
Global bonds	358	182.3	0.402	2.0	3.5
Global equities	452	228.3	0.374	1.6	2.8
Growth and income	978	158.4	0.830	2.5	5.5
Ginnie Mae	182	144.0	0.460	2.4	4.0
Gov't securities	450	131.9	0.549	2.5	4.7
International equities	1267	225.5	0.432	1.9	3.2
Income	218	170.8	0.415	2.2	3.4
Long-term growth	1812	179.4	0.421	2.0	3.1
Tax-free money market	455	62.7	0.440	1.6	3.2
Gov't securities money market	437	59.5	0.611	1.8	4.8
High-quality muni bond	541	137.2	0.624	2.4	4.1
Single-state muni bond	1326	150.3	0.384	1.7	3.6
Taxable money market	541	79.2	0.726	2.0	7.1
High-yield money market	62	160.4	0.408	1.7	3.3
Precious metals	35	256.1	0.399	1.6	3.3
Sector funds	511	200.8	0.364	1.8	2.9
Total return	323	178.2	0.415	1.9	3.3
Utilities	94	182.8	0.359	1.7	3.2
Retail S&P 500 index funds	**82**	**97.1**	**0.677**	**3.1**	**8.2**

The table shows measures of price (fee) dispersion among mutual funds in various asset classes in 2000. The fund data used to calculate these figures is from the Center for Research in Security Prices (CRSP). Roughly 95 percent of the funds in the CRSP database are matched to one of these sectors, which are categorized according to the Investment Company Data, Inc. (now Standard and Poor's Micropal) system. We follow the CRSP convention (also common in the literature) of treating each fund share class in multi-class funds as a separate fund. Multi-class funds are those which have a common manager and portfolio, but have different pricing schemes and asset purchase and redemption rules. Prices are computed from Center for Research on Security Prices (CRSP) data, and are calculated as the fund's annual fees (both management fees and 12b-1 fees if applicable) plus one-seventh of total loads, assuming a mean holding horizon of 7 years, as in Sirri and Tufano [1998]. All prices are expressed in basis points. In 2000, a second ETF started trading, Barclay's iShares S&P 500 Index Fund. See Data Appendix for details.

Table II
Evolution of Retail S&P 500 Index Fund Sector

Variable	Statistic	1995	1996	1997	1998	1999	2000
	N (return data)	23	24	36	54	65	76
Annual gross	Mean	37.43	22.67	33.24	28.95	20.95	-8.63
return	Std. dev.	0.25	1.29	0.19	0.84	0.40	1.23
(percent)	Interquartile range	0.31	0.30	0.20	0.26	0.21	0.32
Std. dev. of	Mean	1.492	3.133	4.574	6.199	3.808	4.932
monthly returns	Std. dev.	0.025	0.038	0.050	0.051	0.115	0.219
(percent)	Interquartile range	0.016	0.023	0.029	0.028	0.024	0.037
	N (price data)	24	33	45	57	68	82
	Minimum	19	18.0	16.0	17.0	17.0	9.45
	25^{th} percentile	43.8	45.0	40.0	46.0	47.1	47.0
	Median	77.5	60.0	70.0	82.0	80.9	72.1
Price	75^{th} percentile	120.5	123.1	136.3	136.3	152.9	144.8
(basis points)	Maximum	206.4	206.4	231.4	231.4	235.4	268.4
	Mean	82.4	80.6	89.8	94.2	104.2	97.1
	Asset-weighted mean	26.8	26.6	26.0	28.9	31.9	32.2
	Std. dev.	50.5	53.1	61.6	60.5	67.1	65.7
	C_1	78.9	77.0	69.9	62.9	59.9	53.9
	C_4	89.0	88.4	85.9	82.1	80.1	77.8
Market shares	Herfindahl	6281	5992	5003	4127	3776	3208
	Low-price decile	86.0	84.9	80.8	77.6	75.5	74.9
	High-Price Quartile	1.4	1.6	1.7	2.5	3.4	4.1

The table shows characteristics of our retail (i.e., non-institutional) S&P 500 index funds sample. As in Table I, we treat each fund share class in multi-class funds as a separate fund. (If we were to count multiple-class funds as single funds, there would be 50 funds in 2000.) We also include exchange-traded funds (ETFs) based on the S&P 500 index. There is only one ETF for most of our sample, Standard & Poor's Depositary Receipts (SPDRs). We report price figures for only 82 funds in 2000 because we lack price information for three funds. Reported returns data is limited to funds reporting returns in every month of given year to avoid comparing figures from funds operating in non-representative portions of the year. C_1 (C_4) is the fraction of sector assets in largest (largest four) funds. The Herfindahl index is the sum of the squared market shares (expressed as percentages). Low-price decile and high-price quartiles are the combined market shares of the lowest-price and highest-price fund quantiles, respectively.

Table III
Search Model with Unequal Sampling Probabilities

Parameters	(A)	(B)	(C)	(D)	(E)
Log(mean search cost)	-6.17	-6.68	-6.58	-6.78	-6.33
	(0.06)	(0.06)	(0.13)	(0.26)	(0.11)
Variance of logged search costs	1.88	2.07	1.79	1.89	1.95
	(0.03)	(0.06)	(0.06)	(0.08)	(0.04)
Mean marginal cost, basis points	4	12	11	16	4
	(1)	(1)	(2)	(2)	(1)
α	2.62	2.43	2.58	2.44	3.06
	(0.04)	(0.03)	(0.03)	(0.09)	(0.04)
γ	-	-	-	-	0.11
					(0.03)
Time trend of mean search cost	-0.38	-0.50	-0.29	-0.51	-0.30
	(0.05)	(0.04)	(0.08)	(0.08)	(0.06)
Time trend of search cost variance	0.20	0.18	0.17	0.18	0.16
	(0.05)	(0.04)	(0.02)	(0.02)	(0.02)
R^2, prices	0.92	0.97	0.82	0.94	0.99
R^2, quantities	0.98	0.83	0.98	0.84	0.98
Median search cost (1996), b.p.	21	12	14	11	18
IQR of search cost range (1996), b.p.	5.9 to 75	3 to 50	4 to 46	3 to 41	4.7 to 67
Median search cost (2000), b.p.	5	2	4	1	5
IQR of search cost range (2000), b.p.	0.7 to 28	0.3 to11	0.8 to 23	0.2 to 9	0.9 to 30

This table shows the results from estimating the homogeneous-product/unequal-sampling-probability specification in Section V.B. Standard errors are in parentheses. b.p. = basis points. The sample consists of all retail S&P 500 funds operating in any year between 1995 and 2000 (inclusive), excluding ten fund-year observations for which we do not have fee or asset data. This leaves a sample of 309 fund-year observations. See text for details.

Key to estimated specifications (see Data Appendix for additional details on market share and price measures):

(A) Asset-based market shares; prices are annual fees plus one-seventh of total loads.
(B) Market shares are based on inflows into fund; prices are annual fees plus one-seventh of total loads.
(C) Asset-based market shares; prices are annual fees only, ignoring loads.
(D) Market shares are based on inflows into fund; prices are annual fees only, ignoring loads.
(E) Data as in model (A), but sampling probability is a function of both age and (logged) total number of funds in management company.

Table IV
Utility Function Estimates and Fund Attribute Summary Statistics

Attribute	Utility Weight, basis points (s.e.)	Mean	Std. Dev.
Constant	93.53 (69.94)	N/A	N/A
Any load dummy	-12.11 (33.74)	0.547	0.499
Rear/deferred load dummy	59.57 (37.13)	0.272	0.446
Exchange-traded fund	199.5* (58.88)	0.023	0.149
Number other share classes	2.726 (9.722)	1.621	1.337
log[no. funds in same mgmt. company]	30.97* (12.67)	4.259	1.215
log[fund age]	99.39 (54.81)	1.393	0.728
Manager tenure (yrs.)	3.578 (11.03)	2.922	2.776
Income + capital gains yield (percent)	-6.552* (3.009)	3.248	3.363
Avg. monthly percent diff. between fund and S&P 500 returns	136.4* (56.69)	-0.026	0.106
Std. dev. of monthly returns (percent)	48.22* (8.101)	4.455	1.293
N	309		
R^2	0.354		
Mean of dependent variable	582.9		

The table shows implied utility weights (β) of fund attributes using indirect utility levels implied by the differentiated-products model in Sections V.C and V.D. Standard errors are clustered by fund. Fund age is instrumented for using summary measures (fund counts and average attribute levels) of two other sets of sector funds: other funds managed by the same company (excluding the own-fund observation), if applicable; and those managed by other companies. Instruments vary by fund-year observation. Averages of all non-age attributes (i.e., those listed in the table above) are used in the instrument set. See text for details.

Table V
Summary of Welfare Changes In Vanguard Monopoly Counterfactual

	(1)	(2)	(3)	(4)	(5) = \$1.22m x [(1)-1]	(6) = (2)x(3)	(7) = (2)x(4)	(8) = (5)+(6)+(7)	(9) = (8)÷(2)
Year	Funds	Assets (\$billion)	Search Savings (basis points)	Product Variety Cost (basis points)	Fixed Costs Savings (\$million)	Savings from Search (\$million)	Product Variety Cost (\$million)	Net Welfare Change (\$million)	Indifference Monopolist Price Change (basis points)
1995	24	22.0	65.1	-11.8	28.1	143.2	-26.0	145.3	66.1
1996	33	39.4	54.4	-13.0	39.0	257.7	-51.2	245.5	62.3
1997	44	70.6	44.7	-17.8	52.5	315.6	-125.7	242.4	34.3
1998	57	118.0	40.2	-23.1	68.3	474.4	-272.6	270.1	22.9
1999	68	174.8	42.5	-25.5	81.7	742.9	-445.7	378.9	21.7
2000	82	163.8	23.6	-30.8	98.8	386.6	-504.5	-19.1	-1.2

This table summarizes the welfare effects of a counterfactual industry structure where a Vanguard 500 Index Fund monopoly is imposed. Details on the calculations are found in the text and in the Computational Appendix.

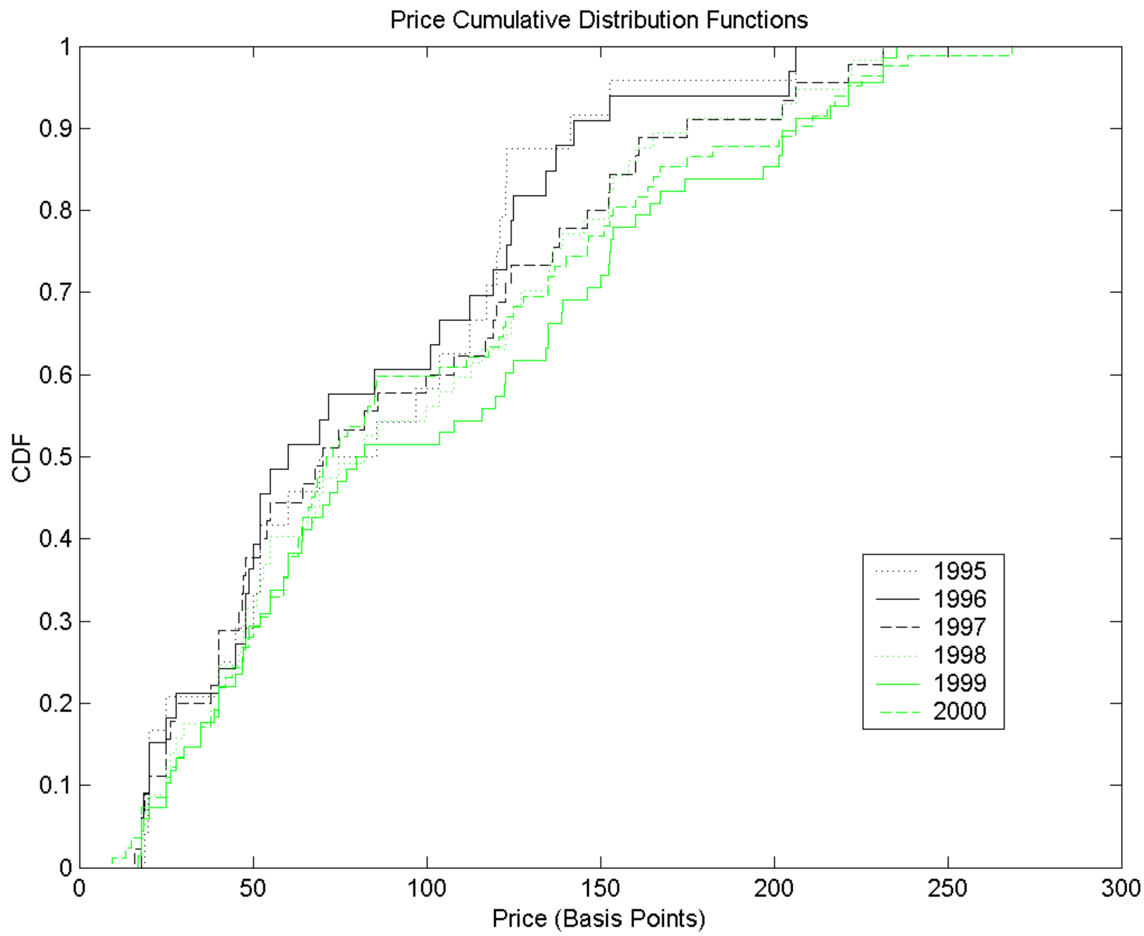

Figure I
The figure shows, by year, c.d.f.s of the retail S&P 500 index fund price (fee) distributions. Prices are computed from Center for Research on Security Prices (CRSP) data, and are calculated as the fund's annual fees plus one-seventh of total loads. All prices are expressed in basis points. See Data Appendix for details.

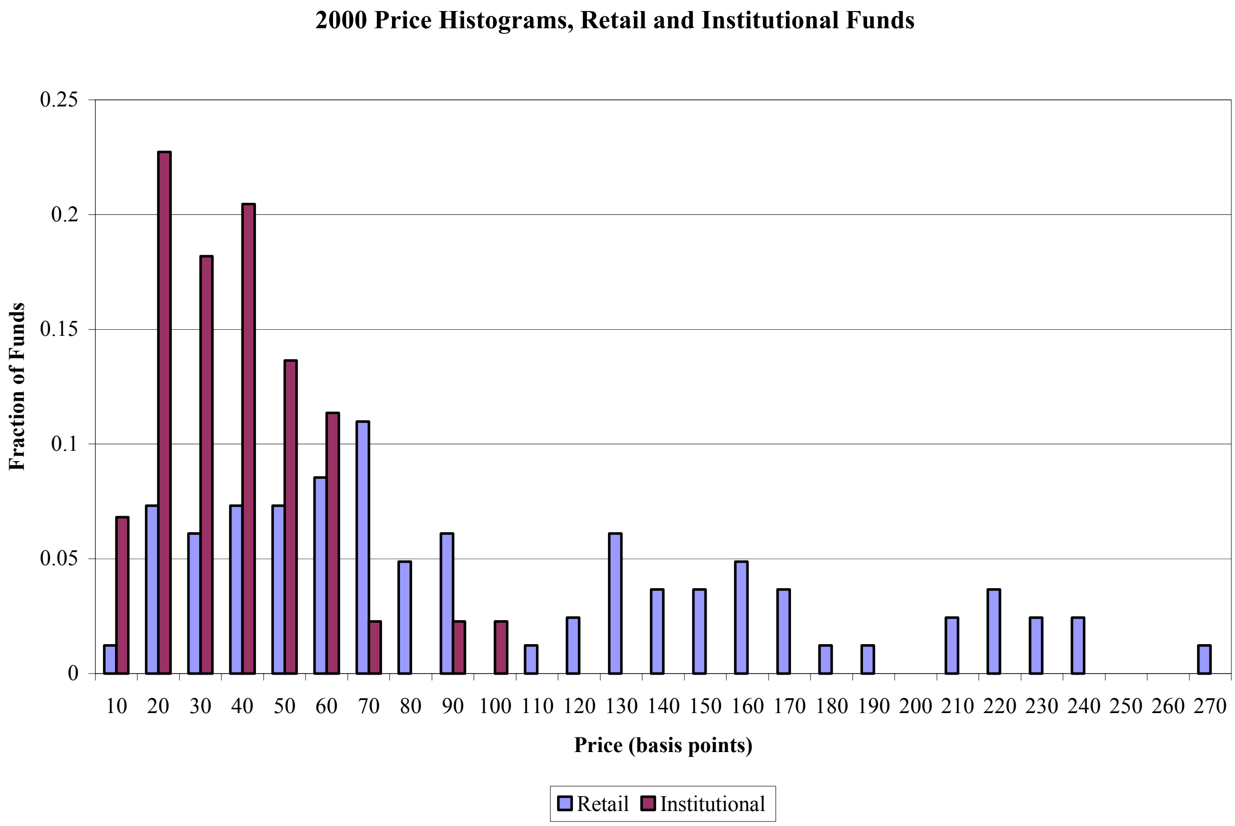

Figure II

The figure shows the price histograms for retail and institutional S&P 500 index funds in 2000. Prices are calculated as annual fees plus one-seventh of total loads. All prices are expressed in basis points.

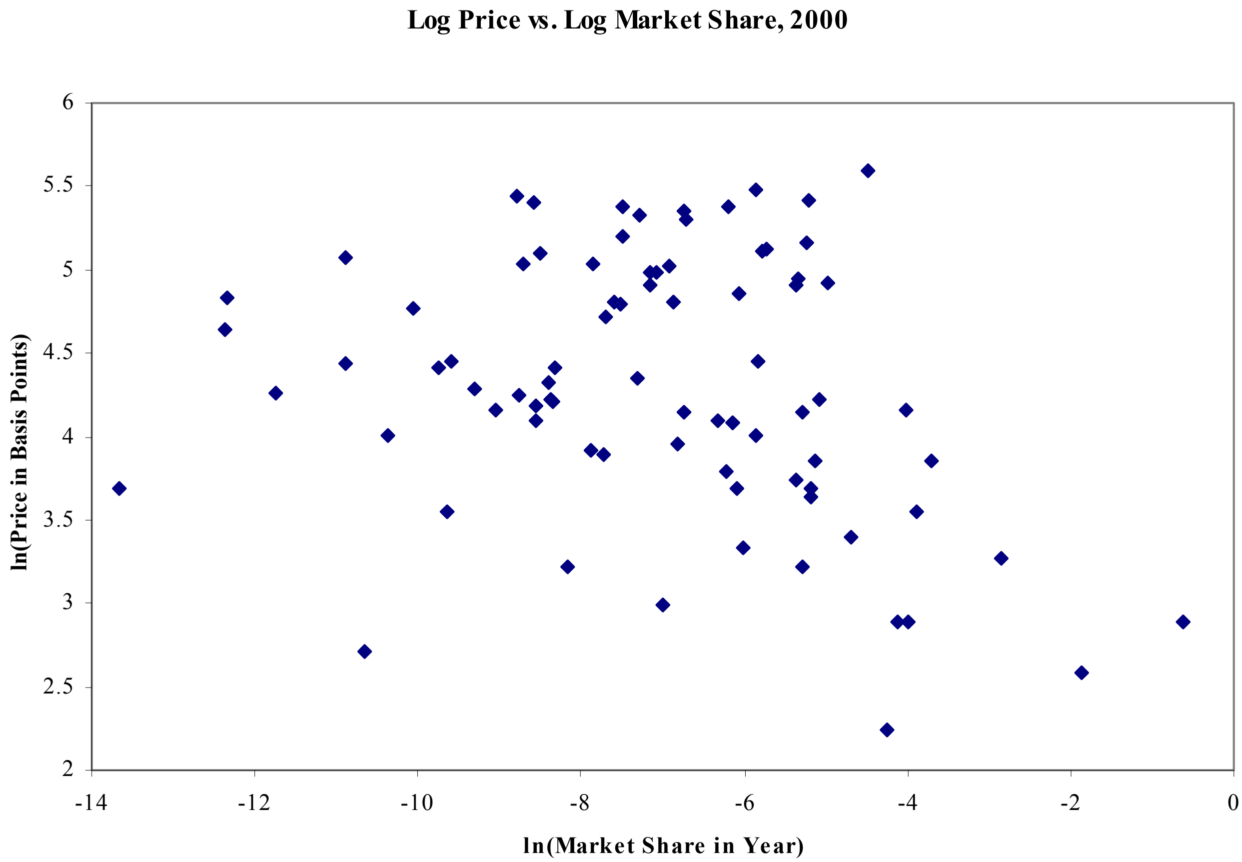

Figure III
The figure plots for retail S&P 500 index funds in 2000 the natural logarithm of market share versus the natural logarithm of fund price. Market shares are based on total fund assets, and price are calculated as annual fees plus one-seventh of total loads (expressed in basis points). See Data Appendix for details.

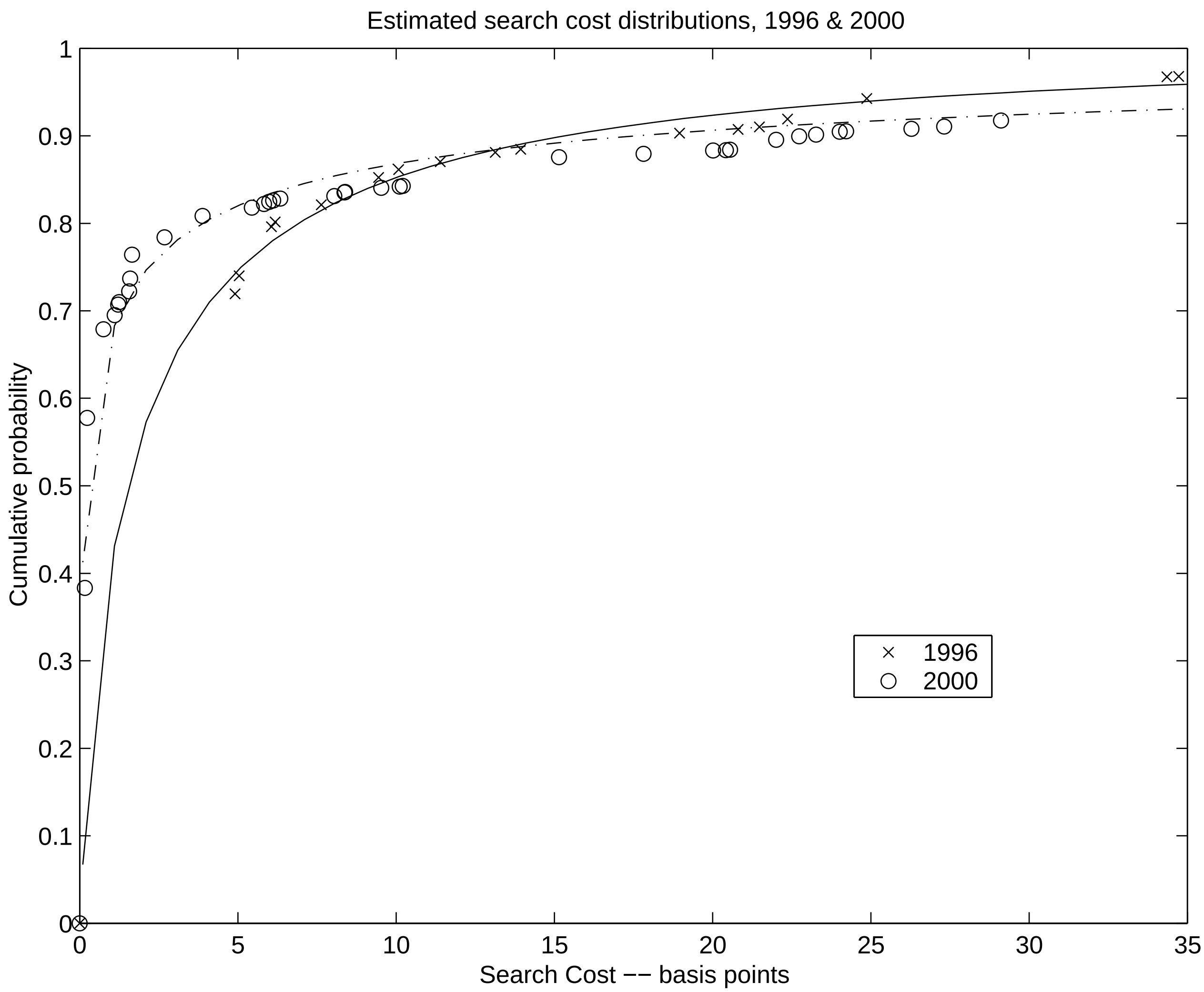

Figure IV
The figure plots the c.d.f.s of the implied search cost distributions from the heterogeneous-funds model in Section V.C. The plotted points are the nonparametrically identified cutoff search costs. The solid (dashed) line is a parametric lognormal distribution fit to the 1996 (2000) nonparametric cutoffs.

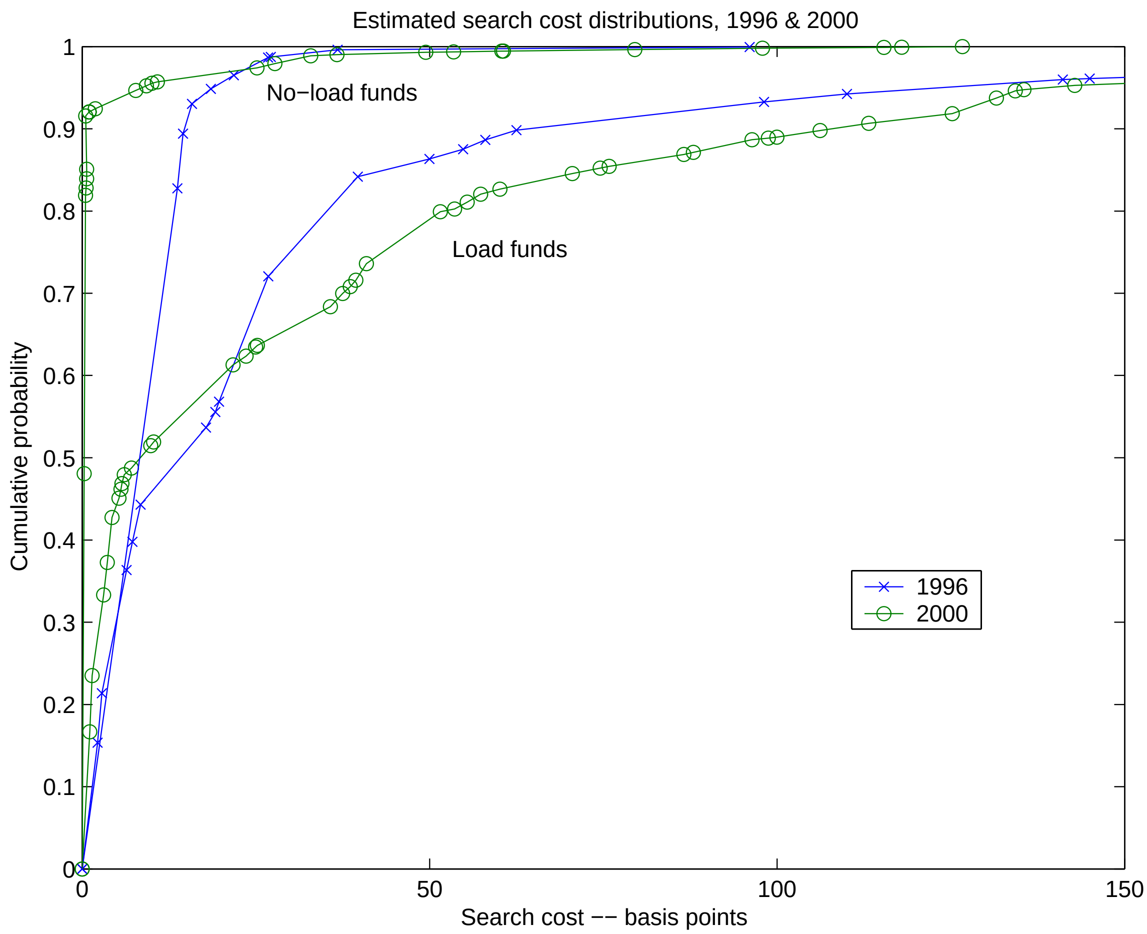

Figure V
The figure plots the search cost distribution c.d.f.s obtained when the heterogeneous-funds model estimation in Figure IV is replicated after dividing investors into load fund buyers or no-load buyers. (See Section V.E for details.) The plotted points are the nonparametrically identified cutoff search costs.