

NBER WORKING PAPER SERIES

ALPHA AND PERFORMANCE MEASUREMENT:
THE EFFECTS OF INVESTOR DISAGREEMENT AND HETEROGENEITY

Wayne E. Ferson
Jerchern Lin

Working Paper 19349
<http://www.nber.org/papers/w19349>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2013

Portions of this work were initially prepared as a keynote address for the 16th Annual Pacific Basin Economics, Accounting and Management Conference and for the 2009 Northern Finance Association Meetings. We are grateful to two anonymous referees, George Aragon, Kerry Back, Stephen Brown, Mendel Fyngenson, Derek Horstmeyer, Juna Joenvaara, Raymond Kan, Min S. Kim, Mark Kritzman, Thierry Post, Sergei Sarkissian, Rick Sias, and to participants in workshops at the University of Arizona, Claremont McKenna College, Erasmus University, the 2011 European Finance Association, Koc University, the 2011 McGill Asset Management Conference, the 2011 Northern Finance Association meetings, the University of New South Wales, the University of Southern California, the University of Sydney, the University of Technology Sydney, the fall 2011 ICI/AIM Investment Center conference at the University of Texas, Tilburg University, the Spring 2011 CQA Conference, the 2011 Wharton Conference on Household and Portfolio Choice and Investment Decisions and the University of Washington for suggestions and discussions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Wayne E. Ferson and Jerchern Lin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Alpha and Performance Measurement: The Effects of Investor Disagreement and Heterogeneity
Wayne E. Ferson and Jerchern Lin
NBER Working Paper No. 19349
August 2013
JEL No. G11

ABSTRACT

The literature has not unambiguously established that a positive alpha, as traditionally measured, means that an investor would want to buy a fund. However, when alpha is defined using the client's marginal utility function, a client faced with a positive alpha would generally want to buy. When markets are incomplete performance measurement is inherently investor specific, and investors will disagree about the attractiveness of a given fund. We provide empirical bounds on the expected disagreement with a traditional alpha and study the cross sectional effects of disagreement and investor heterogeneity on the flow response to past fund alphas. The effects are both economically and statistically significant.

Wayne E. Ferson
Department of Finance
and Business Economics
University of Southern California
3670 Trousdale Parkway Suite 308
Los Angeles, CA 90089-0804
and NBER
ferson@marshall.usc.edu

Jerchern Lin
University of Buffalo
246 Jacobs Management Center
Buffalo NY 14260
jerchern@buffalo.edu

FINANCE RESEARCHERS have an easy familiarity with alpha, the most well-known measure of the abnormal return of an investment. Studies refer to CAPM alpha, three-factor alpha or four-factor alpha, assuming the reader hardly requires a definition. Investment practitioners discuss their strategies in terms of the quest for alpha. Alpha can be active, conditional or portable. The number of investment firms with alpha in their names is truly staggering.

Despite the apparent familiarity with alpha, the current literature too often fails to think rigorously about how alphas can be interpreted. The contributions of this paper are three. First, we offer extensions of existing results that motivate alpha as a guide to investment selection. Given that markets are incomplete and investors have different marginal rates of substitution, the conclusion is that we need to define alpha in a client-specific manner.

Our second contribution is an analysis of how different investor-specific alphas are expected to be from the alphas typically used in research practice. We derive and estimate bounds on the extent to which a client may be expected to disagree with a measure of alpha. Our measure of disagreement integrates out and therefore does not consider differences in beliefs. Even so, we find that disagreement with traditional alphas, based on marginal utility alone, can be similar in importance to the choice of the performance benchmark or to the statistical imprecision in estimates of alphas.

Our third contribution is empirical evidence that investor disagreement and heterogeneity are economically significant in the behavior of fund investors. Funds whose clients' alphas are expected to be larger than a traditional alpha have positive *average disagreement*. We find that such funds experience larger inflows, controlling for the traditional alpha, than funds with low average disagreement. Funds with a large cross-sectional variance of investor disagreement have greater investor *heterogeneity*. We find that funds with more heterogeneity across investors experience lower average flows. These are separate effects from uncertainty due to estimation error in the traditional alpha.

The next section reviews the issues with traditional alphas that motivate our analysis. Section II sets up the problem for investor-specific alphas and Section III provides the main analytical results. Section IV discusses our approach to bounding the extent to which clients may disagree about alpha. Section V discusses the data and Section VI presents the empirical evidence. Section VII concludes and an Appendix provides the proofs.

I. Problems with Traditional Alphas

A Fundamental Question about the use of alphas is: Given a fund with a positive alpha, should an investor buy the fund? While the concept of alpha may be traced in some form back to Cowles (1933), a substantial literature grappled with this question after alpha was developed within the Capital Asset Pricing Model (CAPM, Sharpe, 1964) by Jensen (1968) and others. But this work, with a few exceptions, essentially died out in the late 1980s, leaving the Fundamental Question at best partially resolved.

The literature does offer some hopeful examples that suggest an investor would wish to buy a positive-alpha fund, but there are also many counterexamples. The simplest intuition for the attractiveness of a positive alpha is taught with the CAPM, where a combination of a positive-alpha fund, the market portfolio and cash can "beat the market" in a mean variance sense (higher mean return given the variance). Given an arbitrary (inefficient) benchmark, Dybvig and Ross (1985) show (their Theorem 5), that a positive alpha measured relative to the benchmark implies that buying some of the fund *at the margin*, will result in a higher Sharpe ratio than the benchmark, if the benchmark excess return is positive.

Jobson and Korkie (1982) showed that given an inefficient index, a portfolio with weights proportional to the vector of assets' alphas, premultiplied by an inverse covariance matrix (the optimal orthogonal portfolio), can be combined with the index to

generate a mean variance efficient portfolio. The weight in the optimal orthogonal portfolio for a positive alpha asset can be negative, and Gibbons Ross and Shanken (1989) provide empirical examples where it is. So, even if a positive alpha is attractive at the margin to a mean-variance investor, it might not imply buying a positive alpha fund.¹

There are many counterexamples to the attractiveness of a positive alpha. In some examples performance within the model is neutral but alpha is not zero. Jagannathan and Korajczyk (1986) and Leland (1999) show you can get nonzero CAPM alphas by trading options with no special skill. Ferson and Schadt (1996) show you can record negative alphas when performance is neutral if you don't account for public information. Roll (1978), Dybvig and Ross (1985) and Green (1986) give examples of nearly arbitrary alphas when there is no ability. Goetzmann et al. (2007) show how to produce positive alphas through informationless trading. While the relevant strategies, such as options trading, may be more common among hedge funds and fixed income funds than among the US equity mutual funds examined here², these examples cast doubt on the common presumption that investors would wish to buy a positive alpha fund.

The above examples do not consider differential information. It seems natural to think that a portfolio manager may have better information about returns than the client investor. With differential information the problem of alpha becomes richer (e.g., Admati and Ross, 1985). The portfolio formed by a better-informed manager expands the

¹ In a mean variance setting, the optimal weight need not be of the same sign as the alpha because of correlation among assets. If only a single fund is allowed, in combination with a fixed inefficient benchmark, then the optimal combination of the two does place a positive weight on a positive-alpha fund. Our results generalize this example, as they do not assume that the weights on the other assets in a "market" portfolio are held fixed, do not rely on mean-variance preferences, on a single-period model or on many other restrictions that are imposed in earlier studies. In the simultaneous consideration of multiple assets our results may be applied to a portfolio of the funds.

² For example, Koski and Pontiff (1999) find only limited use of options among US equity style mutual funds.

opportunity set of the less-informed client, so the client would generally like to use the managed portfolio in some way. But a positive abnormal return or alpha is neither necessary nor sufficient to establish a positive value of the manager's information, as observed by Ohlson (1979) in a different setting. The problem here is, the client might wish to short a fund even if it has a positive alpha (Chen and Knez, 1996), or buy a fund even if it has a negative alpha (Glode, 2011). In summary, the existing literature suggests that traditional alphas are not a reliable guide to the attractiveness of an investment.

This paper provides conditions under which the "right" alpha, defined in terms of the client's preferences, provides a reliable buy or sell indication. Our contributions to the literature are both analytical and empirical. First, we derive a well-specified alpha using a multiperiod model under general returns distributions, and show that the client-specific alpha does provide a reliable buy or sell indication. While this is not the first paper to argue that the limitations of alpha can be addressed by defining alpha in a client-specific fashion (e.g. Sharpe (1982), Goetzman et. al., 2007) these earlier studies relied on a single-period model, normality, independence of returns over time and/or specific investor preferences. But, returns are not independently normally distributed over time, and mean-variance preferences are not realistic. Our first contribution is to provide more general results. The cost of this general justification for alpha is that we have to confront investor heterogeneity.

As our second contribution we derive bounds on the extent to which clients are expected to disagree with a traditional measure of alpha and present evidence on the magnitudes. Disagreement presumes that markets are incomplete, because otherwise investors' marginal rates of substitution would agree. Cochrane and Saa-Requejo (2000) study bounds on prices in incomplete markets, of payoffs that are not spanned by a set of benchmark assets. Their pricing bounds, like our bounds on disagreement, are stated in

terms of a maximum Sharpe ratio (Sharpe, 1992) and the volatility of a residual return in a factor model regression. In our bounds the maximum Sharpe ratio is given by the benchmark assets, while in their case it is exogenously specified. Our evidence suggests that investor disagreement can be an effect of similar magnitude to benchmark choice and to the statistical imprecision in estimates of alpha. While benchmark choice and statistical imprecision have attracted considerable attention in the literature, the implications of disagreement and investor heterogeneity have only begun to be explored.

One of the problems with traditional alphas is potential biases from high frequency trading. Goetzmann et al. (2007) propose a "manipulation proof" performance measure (MPPM) that is immune to such biases. Investors may disagree with a traditional alpha because it is manipulated, or because they have different beliefs or preferences. Integrating out the differences in beliefs we find that traditional alphas differ from the MPPM by more than the disagreement with the traditional alphas. We also infer that investors are expected to disagree with the MPPM, at least as much as with traditional alphas, so manipulation is not the main source of disagreement.

The analysis in this paper is related to Glosten and Jagannathan (1994) and Chen and Knez (1996). Glosten and Jagannathan (1994) start with the definition of alpha studied here, based on the stochastic discount factor (SDF) implied by the client's marginal rate of substitution. They assume that clients' SDFs are functions only of a small set of traded benchmark portfolios and options strategies, and focus on the resulting consensus or representative agent valuation. Chen and Knez (1996), following Grinblatt and Titman (1989), focus on the SDF represented by a minimum variance efficient portfolio conditional on the client's information, and allow managers to have better information. They characterize general classes of "admissible" and positive admissible performance

measures, concluding that performance measurement is “essentially arbitrary.”³

Glode (2011) also exploits our key assumption that the covariances of fund returns with investors’ marginal utility may not be captured by standard performance benchmarks. Glode focuses on the consensus pricing of these covariances, while our focus is the disagreement across investors. Finally, our empirical analysis is related to many studies of the relation between fund flows and performance, some of which are discussed in context below.

II. A General Model for a Client’s Alpha

This section revisits alpha using a natural definition based on the stochastic discount factor. The SDF approach was proposed in some form as early as Beja (1971), but became common in asset pricing only after the literature that tried to address the Fundamental Question about alpha had waned. The approach offers new and general insights. These results are based on a multiperiod model and do not require normality. There is no need to rule out timing ability, nor is selectivity information required to be independent of timing information, security-specific or otherwise restricted as in the earlier literature. The client is allowed to have a general consumption response to the managed portfolio. Finally, a mean variance efficient benchmark is not required or used.⁴

Agents make consumption and portfolio choices at each date t , to maximize a time-separable, increasing and concave lifetime utility function, represented as the indirect value function:

³ Our approach uses a “positive admissible” measure. Chen and Knez show that there can be funds that have positive alphas under some positive admissible measures and negative alphas under others. This foreshadows the results here that the “right” alpha is client-specific, and there is likely to be disagreement across investors about alpha.

⁴ Ferson (2012) shows that a mean variance efficient benchmark is only appropriate under quadratic utility.

$$\begin{aligned}
J(W_t, \text{info}) &\equiv \text{Max}_{\{c, x\}} u(C) + E\{ \beta J(W_{t+1}, s_{t+1}) \mid \text{info} \}, \\
\text{s.t. } W_{t+1} &= (W_t - C) x' R_{t+1}, \quad x' \underline{1} = 1,
\end{aligned} \tag{1}$$

where W_t is the wealth at time t , C is the consumption expenditure at time t , x is the portfolio weight, R_{t+1} is the N -vector of gross (i.e., one plus the rate of) returns for the N assets, one of which can be risk-free, and $\underline{1}$ is an N -vector of ones. The K -vector of state variables in the model is s_t at time t and the conditioning information at time t , "info," takes one of two forms. The info is Z_t , representing public information that includes the current values of the state variables s_t and the current risk-free rate if any, when referring to the client. The info is Ω_t when referring to the better-informed manager, assuming that Z_t is contained in Ω_t . The time subscripts are dropped except when needed to avoid ambiguity.⁵

Assuming that the client starts at an interior optimum in the N -asset economy, the first order conditions to the problem imply:

$$E(mR \mid Z) = \underline{1}, \quad \text{with } m = \beta J_w(W_{t+1}) / u_c(C_t), \tag{2}$$

where m is the stochastic discount factor and subscripts denote derivatives. The notation $J_w(W)$ suppresses but allows for the dependence of the value function on the state variables and Z_t .

Consider now presenting the client with a new investment opportunity, the

⁵ There can be another component of wealth; for example nontraded human capital, and that component can imply hedging demands as in Grinblatt and Titman (1989) without affecting the results.

managed portfolio with return $R_p = x(\Omega)'R$ where $x(\Omega)$ is the vector of the informed manager's portfolio weights. We assume that portfolio managers don't invent new securities, just trade the existing ones using better information.⁶ Define alpha for any portfolio R_p as:

$$\alpha_p = E(mR_p | Z) - 1. \quad (3)$$

If the manager has no superior information in the sense that Z includes Ω , then α_p is zero by Equation (2). Let the managed portfolio return be $R_p = v_{t+1}/P_t$, where P_t is the price that the manager offers the client at time t and v_{t+1} is the random value one period later. From the definition of alpha we see that $(1+\alpha_p)P_t = E\{v_{t+1} | Z\}$, so that if alpha is zero the client would find the offer price "fair." A positive alpha suggests a "low" price for the value to the client. It is shown below that this intuition holds when the client's consumption and investments in all assets can change in response to the introduction of the managed portfolio.

III. Addressing the Fundamental Question

When faced with a new investment opportunity R_p with a nonzero alpha, the client will adjust to new optimal consumption and portfolio choices until the alpha is zero at the new optimum.⁷ Consider a situation where we allow the client to adjust current

⁶ It is assumed, as is common in the literature starting with Mayers and Rice (1979), that the manager's trading based on superior information does not affect the market prices of the underlying assets.

⁷ This is a partial normative, not a general equilibrium analysis. Berk and Green (2004) make an argument where equilibrium adjustment comes from the flows of new cash across funds and diseconomies of scale in fund management, which drive informed manager's alphas to zero in equilibrium. Here, the client adjusts with new optimal portfolio and consumption choices. This is similar in spirit to the portfolio optimization procedure in Sharpe (1987), where the client makes discrete changes to arrive at the new optimal portfolio.

consumption and to buy or sell some amount of the manager's fund. The client feels the effects of these decisions in his future wealth, and thus the marginal utilities of current consumption and future wealth change, and the optimal holdings of other assets may change. We assume that the client is a price taker in the markets for investment assets and consumption goods. Let Δ be a reduction in current consumption used to buy the fund, leading to the random wealth at time $t+1$, $W(\Delta) = W_{t+1} + \Delta R_p + (W_t - C)[x(\Delta) - x]'R$, where $x(\Delta)$ is the new optimal portfolio weight vector for the N base assets, normalized to sum to 1.0, and x is the old optimal weight vector.⁸ The appendix proves:

PROPOSITION 1: *Under the assumptions described above, and also assuming regular utility functions to which the mean value theorem applies, and assuming further that the response of the optimal portfolio weights on the original N assets satisfies*

$$[R_{pt+1} + (W_t - C_t) (\partial x(\Delta) / \partial \Delta)'R] [R_{pt+1} + (W_t - C_t) (x(\Delta) - x) / \Delta]'R > 0,$$

Then when confronted with a new investment with an alpha equal to α_p , the client will optimally purchase the discrete amount given by:

$$\Delta = \alpha_p \{u_c / (-u_{cc}^* - Q)\},$$

Where $Q = E\{\beta J_{ww}^ [R_{pt+1} + (W_t - C_t) (\partial x(\Delta) / \partial \Delta)'R] [R_{pt+1} + (W_t - C_t) (x(\Delta) - x) / \Delta]'R\} < 0$ and $u_{cc}^* < 0$.*

The sign of the optimal investment in the new fund is the same as the sign of alpha, so the investor buys (sells) a discrete amount of the fund if alpha is positive (negative). The optimal investment is zero when alpha is zero. The optimal investment is

⁸ The client divides the beginning of period wealth W_t as follows: Δ is invested in the new fund, $C - \Delta$ is consumed and $(W_t - C)$ is invested in the old assets. At $\Delta = 0$, $W(\Delta) = W_{t+1} = (W_t - C)x'R$, and $x(\Delta) = x$. The analysis can accommodate the case where the investor does not change the current consumption, but only the portfolio weights in response to the new investment. In this case the weights $x(\Delta)$ do not sum to 1.0 and $\Delta = W_t(1 - x(\Delta)'1)$. Combinations of these two cases may also be considered.

proportional to alpha and scaled by a term related to "risk tolerance."

The assumption that $[R_{pt+1} + (W_t - C_t)(\partial x(\Delta) / \partial \Delta)' R] [R_{pt+1} + (W_t - C_t)(x(\Delta) - x) / \Delta]' R > 0$ says that the derivatives of the optimal portfolio weights on the N base assets are adequately approximated by the normalized discrete changes. This holds in the limit for small Δ . For discrete Δ the restriction holds when $x(\Delta)$ is well-approximated by a linear function of Δ or when the relative allocation to the original assets does not change very much. The condition should only fail if the allocations to the original assets change drastically when the new fund is introduced.

The client-specific alpha precludes some of the pathologies described above. For example, we assume a monotone increasing utility function, ruling out the quadratic utility that lies at the root of ambiguities discussed in Mayers and Rice (1979) and Verecchia (1980). Trading within the return measurement interval, which Goetzmann et al. (2007) illustrate with several examples, can generate spurious traditional alphas but need not bias the SDF alpha studied here.⁹ Of course, this does not rule out statistical biases in measuring alpha. For example, return smoothing can make it difficult to estimate alpha because the returns of the fund are not accurately measured (e.g. Asness et al. (2001), Getmansky, Lo and Makarov, 2004).

IV. Disagreement and Investor Heterogeneity

The conclusion from the last section is that a client-specific definition of alpha can

⁹ The issue of bias because funds trade within the return measurement period, or interim trading, is raised by Goetzmann, Ingersoll and Ivkovic (2000) and Ferson and Khang (2002), and examined in detail by Goetzmann et al. (2007). However, Ferson, Henry and Kisgen (2006) show that if the right time-aggregated SDF is used the problem is avoided. The definition of alpha here involves the right time-aggregated SDF on the (perhaps heroic) assumption that investors optimize as often as managers trade.

provide a reliable buy/sell signal. But investors will not in general agree about the right alpha, so the same fund may look attractive to one investor but not to another. In this section we define a measure of the expected *disagreement* a client will have with a traditional alpha and derive bounds on its magnitude. We then discuss the implications for mutual fund flows of the expected disagreement and also *heterogeneity*, which we define as the variance of disagreement across investors.

We work with excess returns, $r \equiv R - R_f$, where R_f is a gross short term Treasury return. Thus, Equation (2) implies that $E(mr | Z) = 0$ and equation (3) implies that $\alpha_p = E(mr_p | Z)$. Consider a regression over time of r_p on the "passive" assets $\{r_j\}$ in r :

$$r_p = a_p + \sum_j \beta_j r_j + \varepsilon_p. \quad (5)$$

This is a simple, unconditional regression with constant coefficients and $E(\varepsilon_p) = E(\varepsilon_p r_j) = 0$. Taking the unconditional expectation of the expression for alpha and substituting in Equation (5) for r_p , we obtain:

$$E(\alpha_p) = E(m)a_p + \sum_j \beta_j E(mr_j) + E(m\varepsilon_p). \quad (6)$$

The first term on the right hand side of (6) captures the traditional, unconditional alpha that would be obtained if $\{r_j\}$ were used as the benchmark returns in a factor model.¹⁰ Since $E(mr_j | Z) = 0$ for the passive assets, the second term in Equation (6) is zero. The third term can be expressed as $E(m\varepsilon_p) = \text{Cov}(m, \varepsilon_p) = \rho_{em} \sigma(m) \sigma(\varepsilon_p)$, where ρ_{em} is the correlation

¹⁰ Using the SDF approach, α_p is measured at the beginning of the period, like an asset price. Using the traditional regressions of returns on factors, alpha is measured at the end of the period, like a return. The term $E(m)$ translates between the two dates.

between m and ε_p . If the correlation is zero, the traditional alpha and the expected alpha for the client coincide. Investor disagreement with a measure of alpha arises when the part of the fund return that is not captured by the traditional factors is correlated with the investor's marginal rate of substitution.

Expanding the expectation of the product, mr_j , into the product of the expectations plus the covariance and setting the result to zero we obtain:

$$\sigma(m)/E(m) = (-1/\rho_{mr_j})[E(r_j)/\sigma(r_j)], \quad (7)$$

where ρ_{mr_j} is the correlation between m and r_j and $\sigma(\cdot)$ denotes the standard deviation. Equations (6) and (7) then imply:

$$\begin{aligned} E(\alpha_p)/E(m) - a_p &= \rho_{em} \sigma(\varepsilon_p) [\sigma(m)/E(m)] \\ &= \rho_{em} \sigma(\varepsilon_p) (-1/\rho_{mr_j})E(r_j)/\sigma(r_j). \end{aligned} \quad (8)$$

The term $(-1/\rho_{mr_j})E(r_j)/\sigma(r_j)$ in the second line of Equation (8) is the same for all the passive assets, so it may be replaced with $(-1/\rho_{mr_j^*})E(r_j^*)/\sigma(r_j^*)$, where r_j^* is the portfolio of the passive assets that achieves the maximum Sharpe ratio, SR_{\max} .

A. A Bound on Expected Disagreement

Assuming $SR_{\max} > 0$, equation (8) implies:

$$\begin{aligned} E(\alpha_p)/E(m) - a_p &= (-\rho_{em} / \rho_{mr_j^*}) \sigma(\varepsilon_p) SR_{\max} \\ &\leq \sigma(\varepsilon_p) SR_{\max}, \end{aligned} \quad (9)$$

where we make the assumption that $(\rho_{em} / \rho_{mrj^*}) \leq 1$. Since the return r_j^* maximizes the Sharpe ratio, it maximizes the correlation to m , and this assumption should be innocuous in most cases.¹¹

The first line of Equation (9) defines the *expected disagreement* of the client with a traditional measure of alpha that uses the passive returns r_j as the factors. The second line expresses an upper bound on the expected disagreement that we can estimate using data. The upper bound depends on the volatility of the fund's residuals, $\sigma(\varepsilon_p)$, and on the maximum Sharpe ratio available in the passive returns. When the maximum Sharpe ratio is higher the disagreement is higher because the volatility of a client's marginal rate of substitution is higher. A fund with a low $\sigma(\varepsilon_p)$ is a fund whose return variation is largely captured by the passive benchmark assets. For such a fund there can be little disagreement about its performance.

The maximum Sharpe ratio in Equation (9) should reflect the maximum taken over all of the assets for which the client is at an interior optimum. In empirical practice we can only use small subsets of the possible assets. We consider alternative groups of benchmark assets, in order to assess the sensitivity of the results. Since the maximum Sharpe ratio in the universe of many assets is likely to be larger than on the subsets we examine, our upper bounds on disagreement are conservative from this perspective.

The expected disagreement in (9) depends on the unobserved correlation, ρ_{em} . If that

¹¹ This is an assumption on an unobservable that cannot be verified directly. There could be some mutual fund clients whose marginal utilities are not highly correlated with the passive asset returns and yet are highly correlated with fund residuals. This can occur, as Kerry Back points out in private communication, if both fund payoffs and marginal utilities are highly nonlinear functions of benchmark variables. The assumption can fail in such instances. We then interpret the bounds as applying only to the subset of clients for which the assumption holds. Below we introduce an empirical proxy for state-level client heterogeneity and find correlations consistent with the assumption that $(\rho_{em} / \rho_{mrj^*}) < 1$ in most cases.

correlation is zero there is no disagreement with the traditional alpha. But the evidence below suggests that the correlation is not zero. The second line of Equation (9) suggests that disagreement will be closer to the upper bound for clients with ρ_{mrj^*} far below 1.0. Investors whose marginal utility fluctuations are not well captured by the traditional benchmarks will disagree more with traditional alphas.

It is important to note that we have taken the unconditional expectation of the client's conditional expectation to arrive at these results. The expected alpha should be interpreted as the "econometrician's" expectation of the client's alpha. Clients have different beliefs when they hold different Z s, and their different consumption and portfolio choices lead to different marginal rates of substitution. Applying iterated expectations, we derive implications of investor disagreement that can be represented using the unconditional expectations. Our measure of disagreement is probably conservative on this dimension, because differences in investors' beliefs present another potential source of disagreement about alpha.¹²

B. Disagreement and Fund Flows

Proposition I determines the amount of fund p that investor i would purchase as $\Delta_{ip} = \alpha_{ip} A_i$, where $A_i > 0$. With no restrictions on this latent demand, the net flow to a mutual fund is $F_p = \sum_i \Delta_{ip}$. The flow is positively related to the traditional alpha and a risk-tolerance-weighted average of investors' disagreements with alpha:

¹² Taking the conditional expectation of the client's alpha delivers the expected alpha. The investor's information set is still reflected in the expected alpha to the extent that different information sets lead clients to different consumption and portfolio policies, and thus different marginal rates of substitution. The analysis also applies to conditional moments given public information known by all clients, in which case $E(\cdot)$, ρ and the σ 's refer to these conditional moments.

$$F_p/E(m) = \sum_i A_i [(-\rho_{em} / \rho_{mfj^*}) \sigma(\varepsilon_p) SR_{max}] + (\sum_i A_i) a_p, \quad (10)$$

where the term in brackets $[\cdot]$ is the disagreement from Equation (9). Thus, for a given traditional alpha, a_p , we expect higher flows when the weighted average disagreement is higher. We present empirical proxies below for ρ_{em} and ρ_{mfj^*} , and we proxy for the average disagreement with $ADIS_p$, the unweighted average of the term in brackets. Our first testable prediction is that funds with larger $ADIS_p$ should experience larger flows of investor funds, controlling for the traditional alpha.

We can't observe the risk tolerances A_i that weight the disagreements, so our empirical measures do not address heterogeneity in risk tolerance across investors. The sign of $ADIS_p$ and the risk tolerance weighted average should typically be the same for a given fund, but if there are large covariances between investors' risk tolerances and disagreement and they differ across funds, this presents a source of measurement error that we would expect to weaken our empirical results.¹³

Most of the empirical literature and our evidence below uses the percentage flows to mutual funds. In our model flows are in the units of the numeraire for consumption, and thus arbitrary in scale. The investor's problem makes no reference to the size of a mutual fund. Empirically, the levels of mutual fund dollar flows are highly persistent and may be nonstationary. We therefore follow the literature and use percentage flows in our empirical work, and we include a control variable for the size of the fund.

¹³ Because $A > 0$ and $E\{AD\} = E(A)E(D) + Cov(A,D)$, the covariance between disagreement D and risk tolerance A would have to be large and of the opposite sign as $E(D)$ in order for $E\{AD\}$ to be different in sign from $E(D)$. If $Cov(A,D)$ is similar in magnitude across funds, it is approximately a cross sectional constant in our analysis.

C. Heterogeneity and Fund Flows

We define the *heterogeneity* for a fund p as the variance of the disagreement across investors. In general, the effect of heterogeneity on fund flows is ambiguous. If the cross-sectional variance of the latent demand increases in a mean-preserving spread, for example there could be no effect on the expected flow. However, heterogeneity likely affects fund flows when constraints inhibit investors' demands, even when the latent demand is unaffected by the heterogeneity.

For many reasons, the full range of latent demands will not be expressed in fund flows. For example, investors cannot sell short mutual funds. Chen, Massa and Zhang (2012) describe how this affects flows when investors have different beliefs. Costs like load fees, transactions charges and unrealized capital gains also impede fund flows. Investors' learning about funds and other participation costs (Sirri and Tufano (1998), Huang, Wei and Yan, 2007) also likely impede fund flows. We therefore consider the effects of heterogeneity in the presence of restrictions on the latent demands.¹⁴

For simplicity, assume that the realized flow is truncated, reflecting only the latent demands Δ_{ip} above some lower boundary values, Δ_{iL} . For example, a no-short-selling constraint implies $\Delta_{iL}=0$ for investors not currently holding the fund. In general the truncation point will vary across funds and investors as functions of the characteristics of both funds and investors. In this example heterogeneity can have both a level effect and interaction effects on fund flows.¹⁵

¹⁴ Our definition of alpha is connected to constraints. If we include the fund p among the base assets, with the constraint that the investor may not hold the fund, then the shadow price of the constraint is proportional to the alpha of Proposition I multiplied by the marginal utility of consumption.

¹⁵ Related arguments about constraints, although not stated in terms of our version of heterogeneity, appear in Sirri and Tufano (1998), Jain and Wu (2000), Massa (2003), Nanda et al. (2004), Huang et al. (2007) and Cheng, Massa and Zhang (2012). In the latter, a negative level effect of heterogeneity arises from managers' equilibrium response to

The level effect of heterogeneity on fund flows should be negative for a given average latent demand, because in funds with more heterogeneity more investors' demands will fall below Δ_{iL} and be restricted. This negative level effect is our first testable prediction about heterogeneity.

Heterogeneity should also act to modify the relation between fund flows and performance. A fund with no heterogeneity would have no flow response to performance below Δ_L , and in our model the response would be linear, with a slope determined by risk tolerance, to performance above Δ_L . The flow-performance graph would resemble a hockey stick. With heterogeneity investors will be restricted from responding at different levels of performance, with more investors restricted at lower levels of performance. Fewer investors will be restricted at higher levels of performance, and for very high levels flows should be nearly linear in performance. The flow-performance graph would resemble the sum over many hockey sticks, with different slopes, placed at various points relative to the x-axis.

Our analysis does not model trading costs and taxes, although constraints on latent demands likely depend on these and other costs. Some frictions such as taxes may operate at the higher performance levels. Some costs differ across investors and may imply additional investor heterogeneity that we do not capture. Tax-induced heterogeneity, for example, is likely to be greater in funds that have recently experienced more volatile returns. In our empirical analysis we control for recent fund return volatility and other fund characteristics that are likely to be related to errors in the specification.

V. The Data

A. Fund Returns

We study monthly returns on actively managed US equity mutual funds from January

convexity in the flow-performance relation.

1984 to December 2012, from the CRSP mutual fund database. We remove sector funds, balanced funds and funds other than diversified, US equity funds.¹⁶ We combine share classes for each fund into a total net asset weighted average.

We examine in some experiments, institutional and retail share classes, aggregating the two share classes separately.¹⁷ We remove index funds from the active fund sample, and use them as benchmarks. Index funds are identified by matching fund names with a list of text strings from Busse and Tong (2012, footnote 1). We also check manually that the index funds are correctly identified.

Fama and French (2010) point out a selection bias due to missing returns from about 15% of the funds on CRSP before 1984. Evans (2010) documents that incubated mutual funds overstate average performance due to backfilling. We therefore remove

¹⁶ We select funds as follows. CRSP uses Policy and Wiesenberger objective codes (WB_OBJ) prior to 1993. Strategic Insight Objective codes (SI_OBJ) cover from 1993 to September, 1998 and Lipper Objective codes (Lipper_OBJ) and Lipper Class codes (Lipper_CLASS) are used up to 2008. Most recent funds are classified by Thomson Reuters Objective codes (TR_OBJ). Mid cap funds are coded as GMC in WB_OBJ, MC in Lipper_OBJ, MCCE, MCGE, MCVE in Lipper_CLASS, or MID in TR_OBJ. Small cap funds are coded as SCG in WB_OBJ, MR, SG in Lipper_OBJ, SCCE, SCGE, SCVE in Lipper_CLASS, or SMC in TR_OBJ. Aggressive growth funds are coded as AG, AGG in WB_OBJ, AGG in SI_OBJ, or AGG in TR_OBJ. Growth funds are coded as G, LTG in WB_OBJ, GRO in SI_OBJ, G in Lipper_OBJ, or GRD in TR_OBJ. Growth and income funds are coded as GI, GCI in WB_OBJ, GRI in SI_OBJ, GI in Lipper_OBJ, or GCI in TR_OBJ. Equity income funds are coded as EI, IEQ in WB_OBJ, EI in Lipper_OBJ, EIEI in Lipper_CLASS, or EQI in TR_OBJ. Our “Other” funds category comprises Leverage and short mutual funds, coded as HEDGE in Policy, OI in WB_Obj, OPI in SI_OBJ, CA, DL, DSB, LSE in Lipper_OBJ, or ELCC, SESE in Lipper_CLASS.

¹⁷ Share classes are first identified by fund names and cross checked with the MFLINKS tables in CRSP, which provide Wharton Financial Institution Center Numbers (WFICN) to suggest what funds should be aggregated based on common holdings. Institutional share classes are filtered by the institutional fund flag in CRSP, and fund names are matched

fund returns before the date of fund organization, remove the first year of each fund's returns and remove funds with less than 5 million dollars of assets under management at the end of the previous year. Funds that invest more than 25% in bonds plus cash at the end of previous year are also screened out (30% produces similar results). These screens leave us with a sample of 144 active equity funds in the first year, 1984, growing to 4595 funds in the final year, 2012. We apply the same screens to the index funds as to the active equity funds. This leaves us with 8 index funds in 1984, growing to 1412 in 2012.

B. Benchmark Returns

Since our bounds are relative to a traditional alpha based on a given set of benchmark returns, we include benchmarks that have been common in the literature, and other benchmarks that represent realistic alternative investment choices. We vary the number of benchmarks in the factor model to assess the sensitivity of the results to the number. Our benchmarks are: (1) a broad stock market portfolio, (2) the three Fama and French (1993) factors, (3) a set of six indexes formed from index mutual funds with different strategies, and (4) a set of five exchange traded funds (ETFs).

The proxy for the market portfolio is the CRSP value-weighted index of NYSE stocks. The Fama and French factors are the market excess return, SMB, and HML. SMB and HML measure the excess returns of small caps over big caps and of value stocks over growth stocks, respectively. These data are from Ken French's website¹⁸. We assume that

with "institutional", Inst, Y, I, Class I, Class Y, CL I, and CL Y.

¹⁸ http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

the alphas are zero, from the client's perspective, on the benchmark assets r_j . It seems plausible that the investor would be at an interior optimum relative to his allocation to a broad market index, and thus the alpha of the market index would be zero. The assumption of zero alphas is less plausible for benchmarks like the Fama and French (1993) factors, given the highly tilted small and value stock positions they embed and the short positions they imply.¹⁹

While active funds may have alphas, it seems plausible to assume that investors would agree that index funds and passively-managed exchange traded funds (ETFs) have zero, or nearly zero, alphas. We therefore examine index funds and ETFs as benchmarks. We form six equally-weighted portfolios of index funds, based on their stated benchmarks as reported by Morningstar Direct. The Categories are: S&P500, S&P Midcap, Small cap, Russell, MSCI US and Others.

Exchange traded funds (ETFs) now cover a wide array of asset classes and strategies. We select the following five ETFs based on their passive management, high trading volume and market sector coverage: SPY (S&P 500), MDY (S&P Mid-cap 400), IJR (S&P Small-cap 600), QQQ (Nasdaq 100 Index), and IYR (Dow Jones US Real Estate Index).

¹⁹ If the benchmark assets have nonzero alphas from the client's perspective then the second term of Equation (6) is not zero. Our bounds on disagreement should then include a weighted average of these unobservable client benchmark alphas. The weights are the active fund's betas on the benchmark assets. The weighted average of the alphas of the benchmarks may be close to zero, but our results should be interpreted with caution in view of this issue.

C. Proxies for Disagreement

Investor disagreement differs across funds, according to our model, in proportion to the product $(-\rho_{em} / \rho_{mrj^*}) \sigma(\varepsilon_p)$. We estimate $\sigma(\varepsilon_p)$ from the mutual fund and benchmark returns data, but ρ_{em} and ρ_{mrj^*} are more challenging. We construct proxies for these correlations following Da and Yun (2010), who find that electricity consumption works better than nondurables plus services consumption expenditures in consumption-based asset pricing models. We use annual electricity consumption growth, measured for the 50 states and the District of Columbia during 1984-2012.²⁰ We take the cross-state variation in the time-series correlations between a fund's residual return, ε_p , and the state electricity consumption growths as a proxy for the variation in ρ_{em} across "clients." Of course, the electricity growth correlations do not provide a direct link to marginal utility, so this proxy for the variation in disagreement is crude. Perhaps, better measures can be developed in future work using micro level data.²¹

²⁰ We are grateful to Zhi Da and Hayong Yun for allowing us to use their data.

²¹ We consider two additional proxies for investor heterogeneity based on the various share classes issued by a fund. We separate out "institutional" share classes on the assumption that individuals are more heterogeneous than institutional investors. We also use the number of share classes as a proxy for the degree of heterogeneity in the investor base. Both proxies are correlated with greater heterogeneity using our main measure, but we find that neither proxy is associated with significantly different flows for a given performance measure. In hindsight, this may not be too surprising. Multiple share classes might increase flows because they offer a heterogeneous menu of cost structures, and this can reduce constraints on some investors' latent demands, thus offsetting the greater heterogeneity. Institutional flows are likely determined in large part by considerations outside the scope of our model, such as the special features of 401(k) accounts.

D. Summary Statistics

Table I presents summary statistics for the monthly returns of the equity funds and the benchmark factors. The index funds have mean returns around one half percent per month and standard deviations around five percent. The monthly Sharpe ratios vary between 0.078 and 0.125. The ETFs display more dispersion, with mean returns between 0.28% and 0.87%, standard deviations between 4.3% and 8.0%, and Sharpe ratios between 0.034 and 0.150. With more disperse assets, a benchmark should generate higher maximum Sharpe ratios and higher correlations with a maximum correlation portfolio. These have offsetting effects on the magnitude of disagreement according to Equation (9). The autocorrelations are small for all the benchmarks, with none above 0.15.

The statistics for the mutual fund returns in Panels B and C summarize the cross-section by sorting the individual funds and reporting the values at various fractiles of the cross-sectional distribution. Each column is sorted separately on the statistic shown. The median fund with at least 12 monthly observations returns 0.48% per month more than a Treasury bill over our sample period, and ten percent of the funds have excess returns of 1.2% or more per month. The median standard deviation is 5.0%, but 10 percent of the funds have standard deviations below 3.6% and ten percent are above 6.8%. Ex post, at least, the universe of funds presents a wide range of investment outcomes. Our question is whether investors will disagree about the attractiveness of these investments.

Panel C of Table I presents a comparison sample where only two months of return data are required for inclusion. This shows that requiring 12 months of data, as we do in

the subsequent analyses, has an impact on the distribution of the average returns. This is a form of sample selection-induced survivorship bias. The surviving funds in the left tail do better, and the 12-month survivors have a more compact distribution of average returns. At the one percent left tail the mean return values are -2.53% in the longer-surviving subset, versus -3.41% in the broader sample. The effects on the standard deviations are smaller. Our measures of disagreement and heterogeneity use funds' (residual) standard deviations. Survivor selection seems to slightly inflate the left tail of the standard deviation distribution (1.93% versus 1.57% at the 1% tail). This suggests that our upper bounds on investor disagreement may be slightly too high for the low-volatility funds, in view of survivor selection.

Panel D of Table I summarizes correlations of the electricity growths with the mutual fund residual returns, ρ_{em} . The table summarizes results for the average fund. For the average fund the differences between the maximum and minimum values across the states ranges from 0.09 for the index mutual fund benchmark to 0.19 for the CAPM. There is much more variation in the correlations across funds. The bottom rows present correlations using the standard growth rates of US per capita consumption of nondurables plus services. For most of the models (except the FF3 factors) the correlation is similar to that for the median state.

The electricity correlations present indirect evidence in support of the assumption that $(-\rho_{em} / \rho_{mij}^*) < 1$, which was used to derive the upper bound on disagreement. The ratios of the correlations using the electricity growth rates are below one in every state for three of the benchmark choices, and the ratio is less than one in more than 76% of the states for the

CAPM. The ratios for the aggregate consumption are also below 1.0.

VI. Empirical Results

We first present some statistics about traditional estimates of alpha that we use to inform our interpretation of the bounds on disagreement. Estimates of the bounds are then presented. Finally, we examine the cross-sectional implications of disagreement and heterogeneity for the flow of investor funds.

A. Traditional Alphas

Table II summarizes estimates of alphas using various benchmarks. The alphas are the intercepts time-series regressions of fund excess returns on the benchmark excess returns. (Conditional alphas are discussed in a later section.) The benchmarks are the market portfolio proxy (Mkt), the three Fama-French factors (FF3) and the vectors of index mutual funds (Idx MFs) or ETFs. We also include the “manipulation proof” (MPPM) performance measure of Goetzmann et al. (2007). This is an SDF alpha for a power utility function of the market portfolio return, with a relative risk aversion set equal to 3.0. A similar measure is advocated by Leland (1999), based on the earlier work of Rubinstein (1976). The cross-section is summarized, as in Table I, by sorting the funds on the value of each statistic and presenting the cutoff values for various fractiles. The statistics include the point estimates of alpha, its standard error and the residual standard deviation of the regression.

Distributions of alpha estimates under the null of zero alphas are generated by

simulation in several recent studies. We compare the cross-sectional distribution of the t -ratios in our data to simulations under the null from Fama and French (2010, Table III) for 1984-2006. Consistent with their findings, there is more dispersion in the distribution of performance than expected under the null of zero alphas. In particular, the left tail of the cross-sectional distribution is thicker than is expected if the alphas were zero.

The residual standard deviations are used in our measures of disagreement. The mean residual standard deviation varies from 1.3% using the vector of ETFs or index funds, to 1.9% using the single market index. There is substantial variation in the estimated $\sigma(\epsilon_p)$ across funds: the interquartile range under the CAPM is 1.1-2.3% per month. This suggests variation in the disagreement across funds, as the residual risk is the risk that investors may disagree about in their evaluation of mutual funds.

The statistics in Table II allow rough calculations about the likely magnitudes of investor disagreement. In the CAPM for example, the median residual volatility is 1.6% per month. The monthly maximum Sharpe ratio is about 0.13, suggesting an upper bound on the expected disagreement of about 0.21% per month, or 2.5% per year for the median fund. The median standard error of the CAPM alpha is 0.21% or about 2.5% per year, so the magnitude of the expected disagreement with the CAPM could be on the order of about one standard error of estimation uncertainty.

Panel B of Table II provides another point of comparison. The first column shows the range of traditional alpha estimates across the four benchmark models. For each fund, the range is the difference between the largest and the smallest alpha, and the distribution of the range is summarized across the funds. The median range of alphas across the

models is 0.20% per month, or about 2.4% per year. This suggests that the magnitude of expected investor disagreement could be on the same order of magnitude as the ambiguity associated with the choice of a benchmark model.

The right hand column of Panel A summarizes estimates of the MPPM alphas for mutual funds. The mean and median values are lower than the other alphas, consistent with manipulation that inflates the traditional alphas. The right-hand column of Panel B reports the range of alphas across the four benchmark models and the MPPM. The range is inflated, compared with the range using the four models, illustrating that the MPPM differs substantially from the traditional alphas for many funds. This raises the question of whether disagreement with a traditional alpha could arise from manipulation.

B. Disagreement with Alphas

Table III summarizes measures of investor disagreement with the traditional alphas and the MPPM. Regressions of the excess fund returns on the benchmark excess returns deliver the alphas and the regression residuals, ε_p . We use the electricity growth data to obtain estimates of the average disagreement for each fund. Panel A summarizes the average disagreement, $E[(-\rho_{em} / \rho_{mrj^*}) \sigma(\varepsilon_p) SR_{max}]$, where $E[.]$ is the sample average across the 51 "states" for a given fund p , $-\rho_{em}$ is the correlation between a fund's residual return and a state's electricity consumption growth and ρ_{mrj^*} is the correlation with the maximum correlation portfolio of the benchmarks. The cross-sectional distributions are summarized as before. The upper bounds at the bottom of the panel are calculated as the product $\sigma(\varepsilon_p) SR_{max}$, where $\sigma(\varepsilon_p)$ is the standard deviation of ε_p and SR_{max} is the maximum Sharpe ratio in the benchmark assets.

It is well known that estimates of maximum Sharpe ratios are upwardly biased in finite samples (e.g. Jobson and Korkie, 1982). We adjust the maximum Sharpe ratio

estimates for finite sample bias following Ferson and Siegel (2003).²²

The average disagreement is close to zero for the average fund, but there seems to be substantial heterogeneity across funds. The 10% tail values in the FF3 model, for example, are -0.29% and +0.34% per month, and they are much larger for the CAPM. We exploit this variation across funds in our analysis of the flow-performance relation. The mean and median upper bounds on disagreement, which do not rely on the electricity data, are shown in the bottom rows of Panel A. The mean upper bound varies across the benchmarks from 0.21-0.38% per month, or about 2.5-4.5% per year. Thus, even for the average fund disagreement with the traditional alphas may be substantial.

Table III allows us to refine the rough calculations above for the economic magnitudes of disagreement. The average upper bound is comparable in magnitude to the ambiguity in traditional alphas associated with the choice of benchmark. A representative example is the FF3 benchmark. Here the median upper bound on disagreement is 0.25% per month. In comparison, for the median fund, the range of alphas across the four benchmark models is 0.20% per month in Table II. The average investor disagreement could be comparable in magnitude to the effect of benchmark choice. The average upper bound is also comparable in magnitude to the effect of estimation uncertainty in the traditional alphas. For example, for the median fund the standard error of the CAPM alpha in Table II is 0.21%. In Table III the median upper bound on disagreement relative to the CAPM is also 0.21% per month. Thus, investors may disagree on the performance of the median fund, as measured by the CAPM, by an

²² The adjusted measure is $[(T-N-2)/T] SR_{\max}^2 - N/T$, where N is the number of benchmark assets, T is the length of the time series and SR_{\max}^2 is the maximum likelihood estimate of the maximum squared Sharpe ratio under normality. We do not apply the correction when $N=1$.

amount similar to the standard error of the traditional alpha.²³

Panel B of Table III summarizes measures of disagreement with the MPPM. We take the average disagreement with a traditional alpha, estimated as described above. We subtract from this the difference between the MPPM and the traditional alpha. The result is an indirect measure of average disagreement with the MPPM. As the indirect measure depends on the traditional alpha, we show results for the different benchmark models.

For each model except the ETF benchmark, Panel B shows that the average disagreement with the MPPM for the mean and median funds is larger than the average disagreement with the traditional alphas. This is consistent with manipulation that makes the traditional alphas “too large.” The numbers also suggest that the range of average disagreement with the MPPM is substantially greater than it is for the traditional alphas, with the exception of the ETF benchmark. Of course, the measures of disagreement with the MPPM involve a more complex calculation with more estimated values, and are thus likely to be noisier. However, the figures suggest that investors are likely to disagree with the MPPM, just as they do with the traditional alphas, and possibly to an even greater extent. Thus, manipulation is not likely to be the main source of investor disagreement with traditional alphas. In the next section we compare the information in these measures for the

²³ One might ask if there are any mutual funds whose traditional alphas are so large that they swamp the disagreement, and we would expect all investors to agree that they are positive. Fama and French (2010) find insignificant traditional alphas for net returns, in both the positive and negative tails of the cross-sectional distribution of equity funds in their more recent subperiod. If the largest traditional alphas are insignificantly different from zero, they will not be significantly larger than zero plus our bounds on disagreement. While it is conceivable that most investors could disagree with a traditional alpha in the same direction (e.g., the CAPM alphas for small stocks are too low), this suggests systematic missing factors in the model. Absent such factors there are unlikely to be many funds where all investors agree that the alpha is positive (or negative), consistent with the large dispersion of average disagreement in Panel A of Table III. Thus, investor disagreement is likely to be important in the evaluation of almost all mutual funds.

flow of investor funds.

C. Cross-sectional Implications of Disagreement

Our first prediction is that we should find a positive level effect of average disagreement on fund flows, controlling for a traditional alpha. Mutual fund flows are measured in the usual way as:

$$\text{Flow}_{p,t} = [\text{TNA}_{p,t} - \text{TNA}_{p,t-1} R_{pt}] / \text{TNA}_{p,t-1}, \quad (11)$$

where TNA is the total net assets of the fund and R_{pt} is one plus the rate of return. Fund flows are measured quarterly and the unit of analysis is the fund quarter.

Our proxy for the average disagreement is $\text{ADIS}_p = E[(-\rho_{em} / \rho_{mrf}^*) \sigma(\varepsilon_p) \text{SR}_{\max}]$, as in Table III. However, in exploring the cross-sectional implications of disagreement and heterogeneity, we must be cognizant of the close relation of ADIS_p to the standard error of the traditional alpha estimate for a fund. Both depend on $\sigma(\varepsilon_p)$, the residual volatility of the fund's return. We do not want to confuse the cross sectional effects of disagreement with the effects of imprecision in estimates of the traditional alpha. We therefore orthogonalize ADIS_p to the standard errors. For each period we run a cross-sectional regression of the ADIS_p 's on heteroskedasticity-consistent standard error estimates for funds' alphas, and take the intercept plus residuals of these regressions as the orthogonalized measure, ODIS_p .

Our first flow performance regression addresses the level effect of average disagreement in a pooled panel regression:

$$\text{Flow}_{p,t+1} = b_0 + b_1 a_{p,t} + b_2 \text{ODIS}_{pt} + \text{controls}_t + u_{p,t+1}. \quad (12)$$

The traditional alphas for a given quarter, $a_{p,t}$, are estimated using 36 months of past data (a minimum of 12 months are required) for each fund.

The control variables include the lagged volatility of the fund's return over the previous 36 (minimum of 12) months. This can be motivated by previous empirical work (e.g. Sirri and Tufano, 1998), as a proxy for uncertainty about managerial ability or to soak up heterogeneity in fund investors' tax bases. We include the funds' size (log of TNA), age (the log of 1 + number of months since fund inception) and the expense ratio (plus 1/7 of the load fee, if any). Chen, Goldstein and Jiang (2010) find that funds holding more liquid assets experience stronger flows in relation to performance. We therefore include a simple measure of liquidity: the first order autocorrelation of the fund's reported return over the past 36 (minimum of 12) months, as suggested by the model of Getmansky, Lo and Makarov (2004). Two variables control for fund family membership: a dummy variable for whether the fund is in a family, and an index for the number of funds in the family.

In the left hand columns of Table IV include year dummies in the controls and thus identify the coefficients mainly from the cross-section. The aggregate quarterly flows to the fund sector replace the time dummies in the right hand columns. Because the residuals are likely to be correlated within fund styles, we cluster the standard errors by time and by fund style.²⁴

Panel A of Table IV presents results using $ADIS_p$ as the disagreement measure and Panel B uses the orthogonalized measure, $ODIS_p$. There is a large and significant positive relation between the alphas and the subsequent flows of investor funds, consistent with

²⁴ The style clusters are mid cap, small cap, aggressive growth, growth, growth and income, equity income and other, where the allocation of funds to the strategies are determined as described in footnote 16. This scheme assumes no correlation of the style-level flows over time. We examine the sample correlations and find them to be insignificant.

previous studies. The signs on most of the control variables are consistent with previous work. For example, the coefficient on fund size is negative and significant. Larger mutual funds experience smaller percentage flows for a given performance measure.

Our main focus is the average disagreement, which has a positive coefficient in each flow regression. The t-ratios of the coefficients are larger than two in five of the six examples. The coefficients are similar using the ADIS and the ODIS measures, and they are relatively insensitive to the controls. The coefficients are in the range 2.7 to 4.0. This says that the flow of new money to the average fund is 2.7% to 4% smaller when the average investor thinks that the FF3 alpha is 1% too high. Using the standard deviation of the ADIS variable and the coefficients in Panel A, the typical fund flow is about 1.0-1.6% higher in a given quarter when ADIS is one standard deviation above its mean (using the cross-sectional variance of ADIS, the figure is slightly higher). For ODIS, the flow is about 1.1-1.3% higher when ODIS is one standard deviation above its mean. In comparison, the unconditional mean flow in our sample is 2.9% per quarter (the median is -1.1%), so the disagreement effect appears to be economically significant.

D. Cross-sectional Implications of Heterogeneity

We predict a negative level effect of heterogeneity on fund flows for a given latent demand. We use different performance measures for the latent demand. One is the sum of a traditional alpha and its associated average disagreement, denoted as Δ_p . We proxy for investor heterogeneity using the standard deviation across the 51 states, of the disagreement measure: $HET_p = \text{Std}[(-\rho_{em} / \rho_{mrf}^*) \sigma(\varepsilon_p) SR_{\max}]$, where $\text{Std}[\cdot]$ is the sample standard deviation taken across the states. We orthogonalize the heterogeneity measures to the standard errors of alpha, using the same approach described above. We denote the orthogonalized heterogeneity measure as $OHET_p$.

Table V presents the results of the flow performance regressions focusing on the level effects of heterogeneity. Three sets of columns are shown in each panel, with different subsets of the control variables. The coefficients on the alphas are positive and significant, with magnitudes similar to those in Table IV. The coefficients on Δ_p are slightly smaller, but strongly statistically significant. The coefficients on the MPPM are the smallest, but they are still statistically significant.

Our main focus is the coefficients on OHET, which are strongly significant and negative in each regression. The coefficients on OHET are slightly larger when using the latent demand Δ_p , compared with using the raw alphas or the MPPM. The coefficients vary across the experiments from -9.5 to -13.5. This says that a fund with 1% more heterogeneity in its investor base experiences about 10% less in future fund flows, other things equal.

Using the standard deviation of the OHET variable and the coefficients in Panel A, the typical fund flow is about 3.2.-3.8% lower in a given quarter when OHET is one standard deviation above its mean. When the FF3 alpha is the performance measure in Panel B, the typical flow is about 2.7.-3.2% lower in a given quarter when OHET is one standard deviation above its mean, and when the MPPM is the measure in Panel C, the figures are 2.8-3.4%. These are about twice as large as the disagreement effects in Table IV. Thus, the effect of heterogeneity is likely of great economic significance.²⁵

E. Conditional Models

Conditional performance evaluation, as introduced by Ferson and Schadt (1996), attempts to obtain better measures of manager skill by stripping out the effects of public information. However, Del Guercio and Tkac (2002) find that individual investors do not

²⁵ In a previous version of this paper we study the interaction effects of heterogeneity with regressions and graphs, and find patterns broadly consistent with our discussion of the interaction effects in Section IV.B. These results are available by request.

respond to complex performance measures as much as to simpler measures. We examine a number of conditional models to assess the sensitivity of the disagreement measures. (Tables are available by request.) The first example follows Ferson and Schadt (1996), with constant alphas and time-varying conditional betas, modeled as linear functions of lagged instruments for public information.²⁶

The conditional models generally find larger bounds on disagreement. The effects of the smaller residual standard deviations (which reduce disagreement) are more than offset by the larger maximum Sharpe ratios in models with a larger effective number of benchmarks. For example, in the FF3 factor benchmark, the median estimate of the upper bound is 0.52% in the Ferson and Schadt model, compared with the 0.21% we found in Table III. At the same time the range of the alpha estimates across the benchmarks is slightly narrower, so the disagreement bound now exceeds the range of alphas across the benchmark models at each fractile of the distribution of funds. Disagreement now appears even more important relative to the effects of benchmark choice and to the statistical imprecision of the alpha estimates.

We allow for the possibility of time-varying conditional alphas, modeling them as linear functions of the lagged instruments, following Christopherson et al. (1998).²⁷ We examine the alphas in NBER recession and non recession periods. The CAPM alphas are slightly larger in recession periods, consistent with previous studies, although we do not

²⁶ We use the one-month Treasury bill yield from CRSP, a yield spread between Moody's Baa-rated and Aaa-rated corporate bonds, and yield difference between a constant maturity 10-year Treasury bond and the 3-month Treasury bill and the sum of the dividends paid on the S&P500 index over the past 12 months divided by the level of the index as lagged instruments. The data are from CRSP and DRI Basic Economics.

²⁷ We also examine models with time-varying conditional volatility using a GARCH(1,1) model. If the fund residual is conditionally heteroskedastic, then its volatility, $\sigma(\varepsilon_p | Z)$, varies over time with Z and the expected bound is $E\{\sigma(\varepsilon_p | Z_t) SR_{\max}(Z_t)\}$, allowing for a covariance between the two terms.

find this to be the case for the FF3 model alphas.²⁸ The standard errors of the alpha estimates in the two subsamples are larger and the range of alphas across the models is greater, reflecting the smaller sample sizes. The inferences about the magnitudes of the disagreement are generally similar to those from the constant alpha models.

VII. Conclusions

The ambiguities in the interpretation of alpha as a normative investment signal are largely resolved when alpha is defined relative to the client's preferences; and, assuming incomplete markets, performance evaluation is inherently client specific. In evaluating managed portfolios, one size does not fit all. Investors will not in general agree about alphas, and the same fund will look attractive to one investor but not to another. We evaluate the effects of investor disagreement and heterogeneity on investment performance measurement and find that the effects are statistically and economically significant. This has important implications for the existing literature, for practical investment evaluation and for future research.

We infer that the expected disagreements of investors with traditional alphas are likely comparable in magnitude to the impacts of benchmark choice and estimation errors. Benchmark choice and statistical imprecision have received a great deal of attention in the literature on performance measurement, but investor heterogeneity has hardly begun to be explored. We find that when measures of the average disagreement across investors indicate that traditional alphas are too low, funds enjoy larger flows of new money for a given traditional alpha. When a fund faces greater heterogeneity, defined as the variance of disagreement across investors, it experiences less flow for a given performance level.

²⁸ Studies finding that mutual funds perform better in recessions include Ferson and Qian (2004), Kowsowski (2011), Moskowitz (2000), and Kacperczyk, Niewerburgh and Veldkamp (2012). These results are criticized by De Souza and Lynch (2012) for *ex post* conditioning.

These effects are separate from uncertainty about the true value of the traditional alpha.

While our analysis indicates that investor disagreement and heterogeneity are economically significant, we likely understate the case. We use iterated expectations to integrate out clients' different beliefs about traditional alphas. We do not consider taxes or transaction costs. These are additional sources of potential disagreement about alpha across investors.

The implications of our results for the practical evaluation of investments are important. One client is likely to view the performance of a given fund differently from another client. If the client's life situation is idiosyncratic, he is likely to view the performance of a fund as idiosyncratically different. Our results also suggest an important avenue for new research on investment performance evaluation. If the client-specific nature of alpha is important, then studies should develop clientele-specific measures of fund performance.

Appendix

A. Proof of Proposition 1:

The first order condition for an optimal response Δ that maximizes the lifetime utility implies:

$$-u_c(C_t - \Delta) + E\{\beta J_w(W(\Delta)) [R_{pt+1} + (W_t - C_t) (\partial x(\Delta) / \partial \Delta)' R] \mid Z_t\} = 0. \quad (A1)$$

Assuming regular utility functions, we use the mean value theorem to represent

$$\begin{aligned} u_c(C_t - \Delta) &= u_c(C_t) - u_{cc}^* \Delta \text{ and} \\ J_w(W(\Delta)) &= J_w(W_{t+1}) + J_{ww}^* [W(\Delta) - W_{t+1}], \end{aligned} \quad (A2)$$

where * indicates that the functions are evaluated at points in the intervals $(C_t-\Delta, C_t)$ and $(W(\Delta), W_{t+1})$ respectively. Substituting (A.2) into (A.1) yields:

$$u_c(C_t) - u_{cc}^* \Delta = E\{\beta [J_w(W_t) + J_{ww}^* (W(\Delta) - W_{t+1})] [R_{p,t+1} + (W_t - C_t)(\partial x(\Delta) / \partial \Delta)' R]\}. \quad (A3)$$

Substituting in $(W(\Delta) - W_{t+1}) = \Delta R_p + (W_t - C_t)[x(\Delta) - x]' R$, where $x(\Delta)$ is the new optimal portfolio weight vector for the N base assets, normalized to sum to 1.0, and x is the old optimal weight vector, and using the first order condition $E\{\beta J_w(W_t) R\} = \underline{1} u_c(C)$ and the fact that $x(\Delta)' \underline{1} = 1$ implies $(\partial x(\Delta) / \partial \Delta)' \underline{1} = 0$, and using the definition of α_p , (A.3) reduces to:

$$u_c(C_t) - u_{cc}^* \Delta = (1 + \alpha_p) u_c(C_t) + \Delta Q, \quad (A4)$$

$$Q = E\{\beta J_{ww}^* [R_{pt+1} + (W_t - C_t) (\partial x(\Delta) / \partial \Delta)' R] [R_{pt+1} + (W_t - C_t) (x(\Delta) - x) / \Delta]' R]\}.$$

Solving for the optimal Δ we have:

$$\Delta = \alpha_p \{u_c / (-u_{cc}^* - Q)\}, \quad (A5)$$

and the conditions of the theorem guarantee that $Q < 0$, which establishes the result. QED.

References

- Admati, A. and S.A. Ross, 1985, Measuring performance in a rational expectations equilibrium model, *Journal of Business* 58(11), 11-26.
- Asness, Clifford, Robert Krail, and John Liew, 2001, Do hedge funds hedge? *Journal of Portfolio Management* 28,6-19.
- Beja, Avraham, 1971, The structure of the cost of capital under uncertainty, *Review of Economic Studies* 4, 359-369.
- Berk, Jonathan and R.C. Green, 2004, Mutual Fund flows in Rational Markets, *Journal of Political Economy* 112, 1269-1295.
- Busse, Jeffrey A. and Q. Tong, 2012, Mutual fund industry selection and persistence, *Review of Asset Pricing Studies* 2, 245-274.
- Chen, Qi, Itay Goldstein, and Wei Jiang, 2010, Payoff complementarities and financial fragility: Evidence from mutual fund outflows, *Journal of Financial Economics* 2010, 239-262.
- Chen, Z., and P.J.Knez, 1996, Portfolio performance Measurement: theory and applications, *Review of Financial Studies* 9, 511-556.
- Chen, Si, and Massimo Massa, and Hong Zhang, 2012, Short-sale constraints and the pricing of managerial skills: The case of Mutual Funds, Working Paper, National University of Singapore.
- Christopherson, Jon A., Wayne Ferson, and Debra A. Glassman, 1998, Conditioning manager alpha on economic information: Another look at the persistence of performance, *Review of Financial Studies* 11, 111-142.
- Cochrane, John H. and Jesús Saa-Requejo, 2000, Beyond Arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79-119.
- Cowles, A., 1933, Can stock market forecasters forecast? *Econometrica* 1(3), 309-325.
- Da, Zhi and Hayong Yun, 2010, Electricity consumption and asset prices, working paper, Notre Dame University.
- Del Guercio, Diane, and Paula A. Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis* 37, 523-557.

De Souza, André, and Anthony W. Lynch, 2012, Does mutual fund performance vary over the business cycle? Working Paper, Fordham University.

Dybvig, Philip H., and Ross, Stephen A. 1985. Performance Measurement Using Differential Information and a Security Market Line. *Journal of Finance* 40: 383-99.

Evans, Richard B., 2010, Mutual Fund Incubation, *Journal of Finance* 65(4), 1581-1611.

Fama Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus Skill in the Cross-section of Mutual Fund Returns, *Journal of Finance*, 65, 1915-1948.

Ferson, W., 2012, Investment performance: A review and synthesis, the *Handbook of Economics and Finance*, edited by George Constantinides, Milton Harris and Rene Stulz (forthcoming)

Ferson, Wayne and K. Khang, 2002, Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds, *Journal of Financial Economics* 65, 249-282.

Ferson, W., Tyler Henry and Darren Kisgen, 2006, Evaluating Government Bond Fund Performance with Stochastic Discount Factors, *Review of Financial Studies* 19, 423-456.

Ferson, Wayne, and Meijun Qian, 2004, Conditional performance evaluation revisited, in Research Foundation Monograph of the CFA Institute (formerly, AIMR), ISBN 0-943205-69-7, 84 pages.

Ferson, W., and Rudi Schadt. 1996. "Measuring Fund Strategy and Performance in Changing Economic Conditions." *Journal of Finance*, 51: 425-462.

Ferson, Wayne, and Andrew Siegel, 2003, Stochastic discount factor bounds with conditioning information, *Review of Financial Studies* 16(2), 567-595.

Getmansky, Mila, Andrew W. Lo, and Igor Makarov, 2004, An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529-609.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the Efficiency of a Given Portfolio, *Econometrica* 57, 1121-1152.

Glode, Vincent, 2011, Why mutual funds underperform, *Journal of Financial Economics* 93, 546-559.

Glosten, L. and R. Jagannathan, 1994, A contingent claim approach to portfolio evaluation, *Journal of Empirical Finance* 1, 133-160.

Goetzmann, W., Ingersoll J., Ivkovic Z., 2000, Monthly measurement of daily timers, *Journal of Financial and Quantitative Analysis* 35, 257-290.

Goetzmann, W., Ingersoll, J., Spiegel, M., and Welch, I, 2007, Portfolio performance manipulation and manipulation-proof measures, *Review of Financial Studies* 20, 1503-1546.

Green, R.C., 1986, Benchmark portfolio inefficiency and deviations from the security market line, *Journal of Finance* 41 (2), 295-312.

Grinblatt, Mark and Titman, Sheridan. 1989. Portfolio Performance evaluation: old issues and new insights. *Review of Financial Studies* 2: 393-422.

Huang, Jennifer, Kelsey D. Wei, and Hong Yan, 2007, Participation costs and the sensitivity of fund flows to past performance, *Journal of Finance* 62, 1273-1311.

Jagannathan, Ravi, and Robert Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 59(2), 217-235.

Jain, Prem C., and Joanna Shuang Wu, 2000, Truth in mutual fund advertising: Evidence on Future Performance and Fund Flows, *Journal of Finance* 55, 937-958.

Jensen, M. 1968. "The performance of mutual funds in the period 1945 -1964." *Journal of Finance*, 23: 389-46.

Jobson, J.D., and Bob Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Econometrics* 10(4), 433-466.

Kacperczyk, Marcin T., Stijn Van Niewerburgh, and Laura Veldkamp, 2012, Time-varying fund manager skill, Working Paper, New York University.

Kosowski, R., 2011, Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions, *Quarterly Journal of Finance* 2011, 607-664.

Koski, Jennifer Lynch, and Jeffrey Pontiff, 1999, How are derivatives used? Evidence from the mutual fund industry, *Journal of Finance* 54, 791-816.

Leland, Hayne, 1999, Performance beyond mean-variance: Performance measurement in a nonsymmetric world. *Financial Analysts Journal* 55: 27-36.

- Massa, Massimo, 2003, How do family strategies affect fund performance? When performance-maximization is not the only game in town, *Journal of Financial Economics* 67, 249-315.
- Mayers, D. and E. Rice, 1979, Measuring portfolio performance and the empirical content of asset pricing models, *Journal of Financial Economics* 7(1), 3-28.
- Moskowitz, Tobias J., 2000, Discussion, *Journal of Finance* 55, 1695-1703.
- Nanda, Vikram, Jay Wang, and Lu Zheng, 2004, Family values and the star phenomenon: strategies of mutual fund families, *Review of Financial Studies* 17(3), 667-698.
- Moskowitz, Tobias J., Mutual fund performance: an empirical decomposition into stock-picking talent, style, transaction costs, and expenses: Discussion, *Journal of Finance* 2000 55(4), 1695-1703.
- Ohlson, James A., 1979, Residual (API) Analysis and the private value of information, *Journal of Accounting Research* 17, 506-527.
- Roll, Richard, 1978, Ambiguity when performance is measured by the security market line. *Journal of Finance* 33: 1051-1069.
- Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics and Management Science* 7, 407-425.
- Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-42.
- Sharpe, William F., 1982, Security Codings: Measuring Relative Attractiveness in Perfect and Imperfect Markets, in *Financial Economics: Essays in Honor of Paul Cootner*, (Prentice Hall, Engelwood Cliffs, NJ., 216-229.
- Sharpe, William F., 1987, An algorithm for portfolio improvement, *Advances in Mathematical Programming and Financial Planning*, vol. 1 (Greenwich, CT: JIA Press Inc.)
- Sharpe, William F., 1992, The Sharpe ratio, *Journal of Portfolio Management*, Winter, 7-19.
- Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53,1589-1622.
- Verrechia, R., 1980, The Mayers-Rice conjecture: A counterexample, *Journal of Financial Economics* 8, 87-100.

Table I
Summary Statistics

The data are from January 1984 through December 2012. The benchmark assets include a stock market portfolio (market), the Fama French three factors (FF 3 Factors), six equal-weighted portfolios of index mutual funds (Index MFs) and a set of five ETFs: SPY (large cap), MDY (mid cap), IJR (small cap), QQQ (Nasdaq 100), IYR (Mortgage/Real Estate). Mean is the sample mean, Std Dev is the standard deviation, Min is the minimum, Max is the maximum, AR1 is the first order autocorrelation and SR is the sample Sharpe ratio. All statistics are computed on excess returns measured net of the one month return on a three-month Treasury bill. Columns for the mutual funds are sorted separately on each statistic shown in Panels B and C. Monthly percentage figures are reported in the first four columns. In panel D, ρ_{em} is the correlation between state level electricity consumption growth and the residual from regressing the fund's return on the benchmarks, and ρ_{mrj}^* is the correlation with the maximum correlation portfolio of the benchmark returns.

Panel A: Summary Statistics for Benchmarks

	Mean	Std Dev	Min	Max	AR1	SR
FF 3 Factors						
Market	0.584	4.559	-23.077	12.416	0.087	0.128
SMB	0.068	3.156	-16.390	22.000	-0.034	0.022
HML	0.293	3.070	-12.600	13.840	0.143	0.095
Index MFs						
S&P 500	0.557	4.450	-22.083	15.718	0.046	0.125
S&P MidCap	0.679	4.868	-21.900	16.041	0.124	0.140
SmallCap	0.454	5.861	-20.270	18.922	0.080	0.078
Russell	0.416	5.319	-31.829	13.609	0.136	0.078
MSCI US	0.492	5.256	-32.640	14.052	0.156	0.094
Others	0.559	4.835	-19.073	14.289	0.078	0.116
ETFs						
IJR	0.656	5.717	-19.798	18.071	0.087	0.115
IYR	0.873	6.639	-31.352	29.605	0.073	0.132
MDY	0.786	5.255	-21.584	14.765	0.125	0.150
QQQ	0.275	8.011	-26.599	23.046	0.104	0.034
SPY	0.488	4.329	-16.555	10.914	0.089	0.113

Panel B: Summary Statistics for Mutual Fund Returns with At Least 12 Months' Data

Fractile	Mean	Std Dev	Min	Max	AR1	SR
Bottom 1%	-2.534	1.931	-32.402	3.364	-0.322	-0.326
Bottom 10%	-0.293	3.652	-22.441	7.550	-0.120	-0.055
Bottom 25%	0.169	4.379	-19.184	10.185	-0.036	0.033
Median	0.482	5.054	-14.342	12.255	0.075	0.097
Top 25%	0.820	5.812	-9.077	14.922	0.159	0.175
Top 10%	1.201	6.861	-7.313	19.059	0.232	0.251
Top 1%	2.034	12.595	-3.708	36.050	0.419	0.436

Panel C: Summary Statistics for Mutual Fund Returns with at Least 2 Month's Data

Fractile	Mean	Std Dev	Min	Max	AR1	SR
Bottom 1%	-3.409	1.565	-31.925	0.850	-0.625	-0.510
Bottom 10%	-0.438	3.414	-22.237	6.088	-0.148	-0.082
Bottom 25%	0.140	4.315	-18.842	9.632	-0.044	0.028
Median	0.478	5.017	-13.069	12.005	0.069	0.097
Top 25%	0.845	5.804	-8.603	14.737	0.159	0.181
Top 10%	1.279	6.908	-6.377	18.669	0.236	0.268
Top 1%	2.839	12.734	-0.804	35.833	0.475	0.747

Panel D: Electricity Growth Correlations for the Average Fund

Fractile	CAPM	FF3	Idx MFs	ETFs
Min State ρ_{em}	-0.096	-0.047	-0.026	-0.066
Median State ρ_{em}	0.017	0.017	0.030	0.004
Max State ρ_{em}	0.096	0.079	0.063	0.090
Min State ρ_{mrj^*}	0.001	0.062	0.204	0.335
Median State ρ_{mrj^*}	0.110	0.337	0.664	0.704
Max State ρ_{mrj^*}	0.449	0.560	0.927	0.925
Fraction with $ \rho_{em}/\rho_{mrj^*} < 1$	76.5	100	100	100
Aggregate Cons:				
Median Fund ρ_{em}	0.021	0.081	0.036	0.014
ρ_{mrj^*}	0.190	0.332	0.673	0.581

Table II
Mutual Funds' Alphas and Residual Volatilities Using Various Factor Models

This table summarizes the cross-sectional distribution of alpha estimates, their standard errors, and the volatility of the idiosyncratic residuals for actively managed mutual funds. CAPM denotes the Capital Asset Pricing Model, FF3 denotes the three Fama-French factors, Idx MFs are the six equal-weighted portfolios of index mutual funds, ETFs are a set of five ETFs. The symbols r_p and r_j denote the fund excess returns and the vector of benchmark excess returns, respectively. All figures are in monthly percentage units. The sample period is January, 1984 through December, 2012. The alphas are the intercepts in the following regression:

$$r_p = a_p + B_p' r_j + \varepsilon_p.$$

Std is the White (1980) standard error of the alpha and $\sigma(\varepsilon_p)$ is the standard deviation of the fund residual. The regressions use actively managed mutual funds with at least 12 monthly returns. The last column reports manipulation-proof performance measures (MPPM) proposed by Goetzmann et al. (2007) with a risk aversion coefficient of 3. In each column funds are sorted on the indicated statistic.

Panel A: Summary Statistics for Traditional Mutual Fund Alphas and the MPPM

	CAPM			FF3			MPPM
	Alpha	Std	$\sigma(\varepsilon_p)$	Alpha	Std	$\sigma(\varepsilon_p)$	Alpha
Bottom 1%	-1.271	0.062	0.426	-1.239	0.052	0.358	-4.309
Bottom 10%	-0.452	0.106	0.746	-0.441	0.090	0.636	-0.689
Bottom 25%	-0.247	0.144	1.079	-0.246	0.120	0.887	-0.350
Median	-0.070	0.205	1.608	-0.087	0.165	1.264	-0.105
Top 25%	0.102	0.289	2.301	0.051	0.229	1.818	0.140
Top 10%	0.307	0.411	3.225	0.194	0.339	2.474	0.478
Top 1%	0.869	1.087	6.747	0.630	0.897	4.856	1.131
Mean	-0.079	0.249	1.872	-0.116	0.204	1.475	-0.205

	Idx MFs			ETFs		
	Alpha	Std	$\sigma(\varepsilon_p)$	Alpha	Std	$\sigma(\varepsilon_p)$
Bottom 1%	-1.584	0.039	0.254	-2.063	0.035	0.277
Bottom 10%	-0.456	0.088	0.520	-0.579	0.085	0.559
Bottom 25%	-0.227	0.118	0.731	-0.320	0.111	0.765
Median	-0.059	0.161	1.105	-0.140	0.151	1.089
Top 25%	0.091	0.228	1.673	0.013	0.213	1.555
Top 10%	0.255	0.339	2.322	0.150	0.316	2.124
Top 1%	0.775	0.919	4.699	0.566	0.879	4.517
Mean	-0.119	0.202	1.326	-0.215	0.189	1.296

Panel B: Range of Alphas Across Models

	Four Benchmarks	Five Models Including the MPPM
Bottom 1%	0.027	0.044
Bottom 10%	0.069	0.110
Bottom 25%	0.119	0.176
Median	0.202	0.304
Top 25%	0.358	0.550
Top 10%	0.603	0.967
Top 1%	2.143	4.338
Mean	0.344	0.546

Table III
Investor Disagreement Measures

Measures of disagreement with alphas using various benchmarks are summarized. CAPM uses the market portfolio, FF3 uses the three Fama French factors, Idx MFs uses the the index funds and ETFs uses are a set of five ETFs as the benchmarks. The symbols r_p and r_j denote the fund excess returns and the vector of benchmark excess returns, respectively. The regression is:

$$r_p = a_p + B_p' r_j + \varepsilon_p$$

The upper bounds are calculated as the product $\sigma(\varepsilon_p) SR_{\max}$, where $\sigma(\varepsilon_p)$ is the standard deviation of ε_p and SR_{\max} is the maximum Sharpe ratio in the benchmarks, adjusted for finite sample bias following Ferson and Siegel (2003). The average disagreement is calculated as the average of $(-\rho_{em} / \rho_{mj^*}) \sigma(\varepsilon_p) SR_{\max}$ across the 51 "states" for a given benchmark choice and fund. ρ_{em} is the correlation between state level electricity consumption growth and the residual from regressing the fund's return on the benchmarks, and ρ_{mj^*} is the correlation between the electricity consumption growth and a maximum correlation portfolio of the benchmark returns. Panel B shows disagreement with the manipulation proof performance measure, MPPM. The figures are in monthly percentage units for 1984-2012.

Panel A: Disagreement with Traditional Alphas

Fractile	CAPM	FF3	Idx MFs	ETFs
Bottom 1%	-7.302	-0.697	-0.567	-0.393
Bottom 10%	-2.602	-0.290	-0.222	-0.166
Bottom 25%	-1.141	-0.137	-0.114	-0.094
Median	0.042	0.028	-0.015	0.001
Top 25%	1.255	0.195	0.082	0.098
Top 10%	2.569	0.343	0.178	0.168
Top 1%	6.111	0.893	0.433	0.384
Mean	-0.008	0.035	-0.021	0.002
Upper Bounds:				
Mean	0.240	0.248	0.379	0.210
Median	0.206	0.212	0.316	0.176

Panel B: Disagreement with the MPPM

	CAPM	FF3	Idx MFs	ETFs
Bottom 1%	-6.824	-1.438	-1.439	-1.777
Bottom 10%	-2.218	-0.526	-0.508	-0.625
Bottom 25%	-0.877	-0.206	-0.182	-0.279
Median	0.143	0.036	0.028	-0.039
Top 25%	1.214	0.296	0.220	0.151
Top 10%	2.549	0.651	0.522	0.398
Top 1%	7.209	4.046	3.428	3.807
Mean	0.154	0.135	0.058	-0.015

Table IV
**Quarterly Panel Flow-performance Regressions using Average
 Disagreement and Three-Factor Alphas**

The data cover January, 1984 through December, 2012. The net flow, $\text{Flow}_{p,t+1}$, is defined as the quarter-to-quarter growth in total net assets (TNA) in excess of fund returns. Three-factor alphas, $a_{p,t}$, are estimated using 36 months of past data on excess returns (a minimum of 12 months are required) and the Fama-French factors as follows:

$$r_{p,\tau} = a_{p,t} + \beta_{mkt} r_{mkt,\tau} + \beta_{smb} r_{smb,\tau} + \beta_{hml} r_{hml,\tau} + \zeta_{p,\tau}, \quad \tau = t - 36, \dots, t - 1.$$

We then run the following quarterly pooled panel regressions for 1985-2012:

$$\text{Flow}_{p,t+1} = b_0 + b_1 a_{p,t} + b_2 \text{ADIS}_{p,t} + \text{controls}_t + u_{p,t+1},$$

$$\text{Flow}_{p,t+1} = b_0 + b_1 a_{p,t} + b_2 \text{ODIS}_{p,t} + \text{controls}_t + u_{p,t+1},$$

The proxy for average disagreement is:

$$\text{ADIS}_p = E[(-\rho_{em} / \rho_{mrj}^*) \sigma(\varepsilon_p) \text{SR}_{\max}],$$

where ρ_{em} is the time-series correlation between a fund's residual return and a state's electricity consumption growth and $E[.]$ denotes the cross-sectional average across the 50 states and the District of Columbia. The symbol r_j^* denotes the maximum correlation portfolio in a given set of benchmark returns and $\sigma(\varepsilon_p)$ is fund p 's residual volatility. SR_{\max} is the maximum Sharpe ratio, adjusted for finite sample bias following Ferson and Siegel (2003). We also orthogonalize ADIS_p to the cross section of standard errors of the traditional alphas each quarter, and take the intercept plus residuals of these regressions as the measure ODIS_p . The control variables include the lagged age (the natural logarithm of months since inception, AGE_t), the lagged size (the natural logarithm of TNA, SIZE_t), the expense ratio plus one-seventh of the front-end load (EXPENSE_t), the aggregate flow into the fund style category (SECFLOW_t), the lagged fund total return volatility (TVOL_t) and the lagged first order fund return autocorrelation, estimated over the past 36 months (AR_t). DFAM_t is a dummy variable equal to one if a fund has other funds in the same family, and zero otherwise. NFAM_t indicates the number of funds greater than one in the same family, and is zero if there is only one fund in the family. The year fixed effects are year dummies. The standard errors are clustered by year and fund style.

 Panel A: Fund Flows and ADIS

	Coeff	Std	Coeff	Std	Coeff	Std
a_{pt}	14.298	4.424	12.362	3.433	12.890	3.450
$ADIS_{pt}$	2.965	1.784	3.958	1.820	3.959	1.756
AGE_t	0.093	0.049	0.077	0.046	0.082	0.046
$SIZE_t$	-0.088	0.042	-0.082	0.041	-0.085	0.042
$EXPENSE_t$	-4.691	4.100	-4.120	3.628	-3.661	3.391
$TVOL_t$	-1.621	0.710	-1.832	0.768	-2.004	0.769
AR_t	0.068	0.046	0.130	0.026	0.124	0.012
$SECFLOW_t$			1.224	0.518	1.488	0.807
$DFAM_t$					0.037	0.077
$NFAM_t$					0.001	0.001
Year Fixed Effects	Yes		No		No	

 Panel B: Fund Flows and ODIS

	Coeff	Std	Coeff	Std	Coeff	Std
a_{pt}	14.325	4.431	12.363	3.425	12.860	3.466
$ODIS_{pt}$	3.428	1.712	2.954	1.312	2.695	1.366
AGE_t	0.093	0.049	0.076	0.046	0.080	0.046
$SIZE_t$	-0.088	0.042	-0.081	0.041	-0.085	0.042
$EXPENSE_t$	-4.623	4.117	-4.010	3.663	-3.557	3.427
$TVOL_t$	-1.488	0.712	-1.624	0.752	-1.793	0.760
AR_t	0.066	0.045	0.119	0.038	0.114	0.030
$SECFLOW_t$			1.261	0.497	1.521	0.798
$DFAM_t$					0.039	0.076
$NFAM_t$					0.001	0.001
Year Fixed Effects	Yes		No		No	

Table V
**Flow-performance Regressions using Heterogeneity and Three
 Factor Alphas**

The data cover January, 1984 through December, 2012. The net flow, $\text{Flow}_{p,t+1}$, is defined as the quarter-to-quarter growth in total net assets (TNA) in excess of fund returns. Three-factor alphas, $a_{p,t}$, are estimated using 36 months of past data (a minimum of 12 months are required) and the Fama-French factors as follows:

$$r_{p,\tau} = a_{p,t} + \beta_{mkt} r_{mkt,\tau} + \beta_{smb} r_{smb,\tau} + \beta_{hml} r_{hml,\tau} + \zeta_{p,\tau}, \quad \tau = t-36, \dots, t-1.$$

We then run the following quarterly pooled panel regressions for 1985-2012:

$$\text{Flow}_{p,t+1} = b_0 + b_1 \Delta_{p,t} + b_2 \text{OHET}_{p,t} + \text{controls}_t + u_{p,t+1},$$

$$\text{Flow}_{p,t+1} = b_0 + b_1 a_{p,t} + b_2 \text{OHET}_{p,t} + \text{controls}_t + u_{p,t+1},$$

$$\text{Flow}_{p,t+1} = b_0 + b_1 \text{MPPM}_{p,t} + b_2 \text{OHET}_{p,t} + \text{controls}_t + u_{p,t+1},$$

The proxy for investor heterogeneity is based on: $\text{HET}_p = \text{Std}[(-\rho_{em} / \rho_{mrj}^*) \sigma(\varepsilon_p) \text{SR}_{\max}]$, where ρ_{em} is the time-series correlation between a fund's residual return and a state's electricity consumption growth and $\text{Std}[\cdot]$ denotes the cross-sectional standard deviation across the 50 states and the District of Columbia. The symbol r_j^* denotes the maximum correlation portfolio in a given set of benchmark returns and $\sigma(\varepsilon_p)$ is fund p 's residual volatility. SR_{\max} is the maximum Sharpe ratio, adjusted for finite sample bias following Ferson and Siegel (2003). We orthogonalize HET_p to the cross section of standard errors of the three-factor alphas each quarter, and take the intercept plus residuals of these regressions as the orthogonalized measure, OHET_p . Δ_p is the sum of the three-factor alpha and its average disagreement, $\text{ADIS}_p = E[(-\rho_{em} / \rho_{mrj}^*) \sigma(\varepsilon_p) \text{SR}_{\max}]$. MPPM is the manipulation-proof performance measure.

The control variables include the lagged age (the natural logarithm of months since inception, AGE_t), the lagged size (the natural logarithm of TNA, SIZE_t), the expense ratio plus one-seventh of the front-end load (EXPENSE_t), the aggregate flow into the fund style category (SECFLOW_t), the lagged fund total return volatility estimated over the past 36 months (TVOL_t) and the lagged first order autocorrelation of the fund return estimated over the last 36 months (AR_t). DFAM_t is the dummy equal to one if a fund has other funds in the same family, and zero otherwise. NFAM_t indicates the number of funds greater than one in the same family, and is zero if there is only one fund in a family. The year fixed effects are year dummies. The standard errors are clustered by year and fund style.

Panel A: Fund Flows, OHET and Latent Demands

	Coeff	Std	Coeff	Std	Coeff	Std
Δ_{pt}	10.179	3.846	8.194	2.717	8.311	2.801
OHET _{pt}	-13.546	4.784	-11.462	3.741	-11.499	3.642
AGE _t	0.111	0.074	0.099	0.070	0.108	0.070
SIZE _t	-0.124	0.049	-0.120	0.047	-0.128	0.048
EXPENSE _t	-6.470	4.296	-6.416	4.175	-5.594	3.938
TVOL _t	-3.003	1.373	-2.944	1.184	-2.908	1.173
AR _t	0.164	0.111	0.180	0.116	0.145	0.106
SECFLOW _t			1.608	1.133	1.847	1.263
DFAM _t					0.009	0.101
NFAM _t					0.003	0.001
Year Fixed Effects	Yes		No		No	

Panel B: Fund Flows, OHET and Three-factor Alphas

	Coeff	Std	Coeff	Std	Coeff	Std
a_{pt}	15.358	5.281	11.855	3.433	12.014	3.581
OHET _{pt}	-11.269	3.973	-9.549	3.117	-9.553	3.020
AGE _t	0.115	0.074	0.100	0.070	0.109	0.069
SIZE _t	-0.126	0.049	-0.121	0.047	-0.129	0.048
EXPENSE _t	-6.218	4.295	-6.058	4.103	-5.219	3.852
TVOL _t	-2.235	1.046	-2.567	1.023	-2.528	1.008
AR _t	0.163	0.105	0.162	0.110	0.127	0.100
SECFLOW _t			1.666	1.145	1.905	1.278
DFAM _t					0.017	0.102
NFAM _t					0.003	0.001
Year Fixed Effects	Yes		No		No	

 Panel C: Fund Flows, OHET and the MPPM

	Coeff	Std	Coeff	Std	Coeff	Std
MPPM _{pt}	7.928	3.156	6.010	2.692	6.375	2.833
OHET _{pt}	-12.165	4.753	-10.093	3.387	-9.923	3.224
AGE _t	0.118	0.077	0.103	0.074	0.114	0.073
SIZE _t	-0.129	0.051	-0.124	0.050	-0.133	0.051
EXPENSE _t	-7.196	4.732	-6.692	4.451	-5.824	4.151
TVOL _t	-0.316	1.143	-1.705	0.983	-1.574	0.937
AR _t	0.119	0.096	0.117	0.104	0.078	0.094
SECFLOW _t			1.495	1.002	1.726	1.129
DFAM _t					0.007	0.105
NFAM _t					0.003	0.002
Year Fixed Effect	Yes		No		No	
