Essays on Mutual Fund Performance Evaluation with Clientele Effects

Thèse

Manel Kammoun

Doctorat en sciences de l'administration-Finance et Assurance

Philosophiae Doctor (Ph.D.)

Québec, Canada

Résumé

Cette thèse étudie la performance des fonds mutuels du point de vue de leurs clientèles les plus favorables. Elle comporte trois essais dans lesquels nous développons et adaptons une approche de mesure de performance qui considère le désaccord entre investisseurs et les effets de clientèle pour répondre à trois questions de recherche.

Dans le premier essai, nous étudions le désaccord entre investisseurs et les effets de clientèle dans l'évaluation de performance en développant une mesure pour les plus favorables clientèles des fonds. La mesure est une borne supérieure de performance dans un marché incomplet sous conditions de la loi d'un seul prix et d'absence de bonnes affaires que sont les investissements aux ratios de Sharpe déraisonnablement élevés. Nous montrons que considérer le point de vue des clientèles les plus favorables résulte en une performance généralement positive. Le désaccord total mesuré par la différence entre les bornes supérieure et inférieure de performance est économiquement et statistiquement significatif.

Dans le deuxième essai, nous diagnostiquons les mesures de performance standards en comparant leurs alphas avec celui des plus favorables clientèles. Les résultats montrent que les modèles linéaires inconditionnels, leurs versions conditionnelles et la mesure basée sur la loi d'un seul prix donnent des performances sévères mais admissibles. Les modèles de consommation ont un problème d'inadmissibilité. La mesure de performance à l'abri de manipulation génère des alphas sensibles au choix du paramètre d'aversion au risque.

Dans le troisième essai, nous proposons une mesure de performance spécifique aux clientèles basée sur les préférences de style des investisseurs dans les fonds mutuels. Considérant le désaccord de performance et exploitant mieux les données de classifications, nous investiguons huit mesures représentant des clientèles ayant des préférences favorables aux styles d'actions basés sur la taille et la valeur. Nous trouvons que les fonds classés selon la taille et la valeur ont des performances moyennes neutres ou positives lorsqu'évalués avec leur mesure spécifique aux clientèles appropriée. La performance des autres fonds est sensible aux clientèles. Les résultats supportent un rôle significatif des clientèles de style en évaluation de performance.

Abstract

This thesis studies the performance evaluation of mutual funds from the point of view of their most favorable clienteles. It contains three essays in which we develop and adapt a performance measurement approach that accounts for investor disagreement and clientele effects to answer three research questions.

In the first essay, we investigate investor disagreement and clientele effects in performance evaluation by developing a measure that considers the best potential clienteles of mutual funds. The measure is an upper performance bound in an incomplete market under the law-of-one-price condition and a no-good-deal condition that rules out investment opportunities with unreasonably high Sharpe ratios. We find that considering investor disagreement and focusing on the best potential clienteles lead to a generally positive performance for mutual funds. The total disagreement measured by the difference between upper and lower performance bounds is economically and statistically significant.

In the second essay, we diagnose the validity of standard performance measures by comparing their alphas with the alpha from a performance measure that evaluates mutual funds from the point of view of their most favorable investors. The results show that unconditional linear factor models, their conditional versions and the law-of-one price measure give severe but admissible evaluations of fund performance. Consumption-based models suffer from an inadmissibility problem. The manipulation proof performance measure generates alphas that are sensitive to the choice of risk aversion parameter.

In the third essay, we propose a clientele-specific performance evaluation based on the style preferences of mutual fund investors. Considering performance disagreement and better exploiting style classification data, we investigate eight measures to represent clienteles with favorable preferences for size and value equity styles. We find that funds assigned to size and value styles have neutral to positive average alphas when evaluated with their appropriate clientele-specific measure. The performance of the other funds is sensitive to the clienteles. Our findings support a significant role for style clienteles in performance evaluation.

Table of content

List of tables

List of figures

List of appendices

List of abbreviations

1-SC: One-style clientele

2-SC: Two-style clientele

BC: Best Clientele

BSW: Barras, Scailled and Wermers (2010)

CAPM: Capital Asset Pricing Model

CARHART: Carhart Model

CCAPM: Conditional Capital Asset Pricing Model

CCARHART: Conditional Carhart Model

CFF: Conditional Fama-French Model

FC: Ferson and Chen (2015)

FF: Fama-French Model

GMM: Generalized Method of Moments

LCG: Large-cap growth

LCV: Large-cap value

LOP: Law-of-one-price

MPPM: Manipulation Proof Performance Measure

SCG: Small-cap growth

SCV: Small-cap value

SDF: Stochastic Discount Factor

Dédicace

À l'âme de ma mère

Remerciements

Je tiens à exprimer mes plus vifs remerciements à mon directeur Stéphane Chrétien, professeur titulaire à l'Université Laval, pour son attention et sa disponibilité malgré ses engagements. Sa rigueur scientifique ainsi que ses directives tout au long de cette thèse m'ont permis de réaliser mon rêve d'être doctorante. Il m'a appris ce que précision et rigueur voulaient dire en recherche. Je le remercie aussi pour son soutien moral et ses qualités humaines d'écoute et de compréhension.

Je remercie aussi les membres de jury, Marie-Claude Beaulieu, directrice du Département de finance, assurance et immobilier de l'Université Laval et Gabriel Power, professeur agrégé à l'Université Laval, pour le temps et l'effort qui ont consacré pour évaluer ce travail. J'exprime ma gratitude à Wayne Ferson, « Ivadelle and Theodore Johnson Chair in Banking and Finance et Professor of Finance and Business Economics à l'University of Southern California », qui a accepté d'être examinateur.

J'adresse toute ma gratitude à mes parents et mes frères qui ont su me supporter, me motiver et m'encourager pendant toutes ces années de travail ardu. Quoi qu'ils aient été dans un autre continent, ils ont été toujours là pour me soutenir dans les moments difficiles. Particulièrement, j'adresse un retentissant hommage à ma mère qui a été toujours la source de ma réussite et ma meilleure confidente. Merci Maman d'être mon pilier fondateur! Paix à ton âme!

Mes remerciements s'adressent également à mon merveilleux époux d'avoir été à mes côtés et de me donner la force et le courage pour faire aboutir ce rêve. Merci d'être très patient et compréhensif durant ces années. Je ne pourrai pas mesurer ton apport dans l'accomplissement cette thèse.

Cette thèse n'aurait pu être réalisée sans le soutien financier de différents organismes : le Ministère de l'Enseignement Supérieur et de la Recherche Scientifique de la Tunisie, la Chaire Groupe Investors en planification financière, le Département de finance, assurance et immobilier de la Faculté des sciences de l'administration de l'Université Laval, l'Institut de finance mathématique de Montréal, CIRPÉE-Laval, la

Avant-Propos

Cette thèse comporte cinq chapitres. Le premier chapitre consiste en une introduction générale qui présente la revue de la littérature ainsi que les questions de recherche, objectifs et contributions de la thèse. Les chapitres 2, 3 et 4 font l'objet de trois articles scientifiques. Le premier article inclut dans de cette thèse (chapitre 2), intitulé « Mutual Fund Performance Evaluation and Best Clientele », est révisé et resoumis au « Journal of Financial and Quantitative Analysis ». Le dernier chapitre présente la conclusion générale.

Les trois articles inclus dans cette thèse ont été rédigés en anglais. Ils sont écrits en collaboration avec mon directeur de thèse Stéphane Chrétien, professeur titulaire au Département de finance, assurance et immobilier de l'Université Laval. Je suis l'auteure principale et la responsable de la collecte et des traitements des données ainsi que des résultats.

Les résultats des articles 1 et 2 (chapitres 2 et 3) ont été présentés dans différents congrès : 2014 Mathematical Finance Days à Montréal, 2014 European Financial Management Association à Rome, 2014 World Finance Conference à Venise, 2014 Financial Management Association Annual Meeting à Nashville, Tunisian Society of Financial Studies 2014 Finance Conference à Sousse, et Congrès 2015 de la Société canadienne de science économique à Montréal. Ces articles ont aussi fait l'objet de séminaires départementaux à l'École des sciences de la gestion de l'Université de Québec à Montréal, au « Ted Rogers School of Management » de « Ryerson University », à la Faculté des sciences de l'administration de l'Université Laval et à l'Université du Québec en Outaouais.

1   Motivation and Research Questions

1.1   Introduction

As it did in the 1980s and 1990s, the U.S. mutual fund industry continues to grow significantly in the new millennium, from total net assets of $6,846 billion in December 1999 to $15,852 billion in December 2014[1]. This growth occurs despite difficult financial market conditions and in light of several studies, starting with Jensen (1968), and including Fama and French (2010) and Barras, Scaillet and Wermers (2010) more recently, that document negative performance for mutual funds after considering trading costs and expenses. These conflicting results between the underperformance of mutual funds and their growing industry continue to be one of the most longstanding issues in academic finance, with Gruber (1996) calling this contradiction a « puzzle ».

Recently, Ferson (2010) and Ferson and Lin (2014) outline an area of research that could potentially provide a resolution to this puzzle. Mutual funds cater to specific clienteles through their management style and other investment choices. In incomplete markets, investors disagree about the attractiveness of a fund, and the fund value for its targeted clienteles is likely to be higher than the values for other groups of investors. As a positive evaluation for the targeted clienteles could justify the continued popularity of the mutual fund industry, Ferson (2010, p. 229) emphasizes that a challenge for future research is « to identify and characterize meaningful investor clienteles and to develop performance measures specific to the clienteles ». Analyzing generally the effect of disagreement and heterogeneous preferences on mutual fund performance, Ferson and Lin (2014) argue that such effects can be similar in importance to the widely documented effects of the benchmark choice problem and statistical imprecision in estimates of alpha.

This thesis includes three essays on the performance evaluation of mutual funds that further explore the arguments of Ferson (2010) and Ferson and Lin (2014). Our main distinctive feature is that we study the performance from investors who are favorable to

---

[1]See p. 9 of the 2015 Investment Company Fact Books (55th edition), published by the Investment Company Institute.

mutual funds, in the sense that they value the funds at an upper bound in a setup where the market is incomplete, preferences are heterogeneous and investments that are good deals are ruled out. For performance evaluation purpose, such setup avoids relying on the point of view of representative investors, as most of the literature has done, and instead concentrates on investors who are potentially the best clienteles of the funds.

As the essays in this thesis are written as stand-alone papers, the general goal of this chapter is to provide an overview of their common themes and establish relations between them. Hence, in this chapter, we first provide a selective review of the mutual fund performance evaluation literature to introduce the main issues motivating this thesis. Second, we present an overview of the research questions, objectives and contributions of the three essays in this thesis. Third, we briefly describe the new performance measurement approach developed and implemented in this thesis to account for most favorable investors. Finally, we conclude by stating how the rest of the thesis is structured.

1.2   Literature Review on the Performance Evaluation of Mutual Funds

1.2.1   Performance Evaluation Approaches

Investors face the challenge of selecting the best mutual funds to invest in. Portfolio performance measures can help them to evaluate investment choices available in the market. Starting with the classic CAPM-based measures of Jensen (1968), Sharpe (1966), Treynor (1965) and Treynor and Mazuy (1966), many performance measures have been developed, with more than 100 ways compiled by Cogneau and Hübner (2009a, b).

The most common measures are based on linear multi-factor models and consider the regression intercept (generally called alpha) as performance value (Lehmann and Modest (1987), Sharpe (1992), Elton, Gruber, Das and Hlavka (1993), etc.). Another prevalent measure use portfolio holdings to infer the manager's ability for market timing and asset selectivity (Grinblatt and Titman (1989), Grinblatt and Titman (1993), Daniel, Grinblatt, Titman and Wemers (1997), etc.). Conditional measures account for public information in evaluating performance (Ferson and Schadt (1996), Chen and Knez (1996), Christopherson, Ferson and Glassman (1998), Ferson and Khang (2002), etc.).

Measures based on the stochastic discount factor (SDF) approach compute SDF alpha as the expected product of the SDF and the portfolio return minus one (Glosten and Jagannathan (1994), Chen and Knez (1996), Dahlquist and Söderland (1999), Farnsworth, Ferson, Jackson and Todd (2002), Ferson, Henry and Kisgen (2006), etc.). As a final example, Bayesian measures use prior beliefs on the manager's ability to outperform the benchmark or the validity of an asset pricing model (Baks, Metrick and Wachter (2001), Pástor and Stambaugh (2002a, b), etc.) in performance evaluation.

Despite the abundance of measures, and the even greater number of studies applying them, empirical results from the literature are difficult to reconcile on many issues. The remainder of this review examines issues that are the most important for this thesis: value added by active management, benchmark choice problem, comparison of performance measures with simulations, heterogeneous preferences of mutual fund investors, mutual fund styles and mutual fund data biases.

1.2.2   Value Added by Active Management

A debate exists on whether active management adds value. Most but not all of the empirical literature concludes that active management does not generate better performance than passive management on average. For example, the classic study of Jensen (1968) finds that active management is unable to outperform the benchmark, concluding that investing in index funds may be better. However, if this result is true, why do investors pay to invest in active management? To answer this question, a more recent literature re-examines the issue of whether or not active management adds value.

Gruber (1996) investigates reasons that motivate investors to buy actively-managed mutual funds, even if they could achieve better performance by investing in index funds. He suggests that some skilled managers can forecast future performance from past performance and that their management ability may not be included in the prices. Baks, Metrick and Wachter (2001) find that, if their priors include past returns, investors would choose actively-managed mutual funds with positive historical alphas. They also argue that it is relevant for investors to base their investment decisions on their prior beliefs about the managers' ability. Pástor and Stambaugh (2002b) conclude that both prior beliefs on the

manager's ability and the selection of the evaluation model may affect the choice of optimal investment. Berk and Green (2004) find that about 80% of managers add value with their active strategies. However, this value is absorbed by management expenses charged to investors. Similarly, Berk (2005) argues that active management adds value, but fund managers capture it for themselves through expenses and fees.

Kacperczyk, Sialm and Zheng (2005) find that managers who concentrate their investments in industries on which they hold information achieve better performance. They document that the superior performance of concentrated mutual funds is mainly due to their stock selection ability. Kosowski, Timmermann, Wermers and White (2006) demonstrate that a sizable minority of growth-oriented fund managers possess skill and earn abnormal returns. Cremers and Petajisto (2009) argue that high-activity managed funds outperform their benchmarks both before and after expenses and fees. However, low-activity managed funds underperform their benchmarks after expenses and fees.

French (2008) compares the cost of active management versus the cost of passive management. He argues that, by choosing passive management, investors could have increased their average annual return by 0.67% between 1980 and 2006. Fama and French (2010) show that few actively managed funds are able to cover their costs. They find that mutual funds perform better when they consider gross returns versus net returns, suggesting that costs are too high. Controlling for false discoveries, Barras, Scaillet and Wermers (2010) argue that 75% of mutual funds exhibit zero alpha. The proportion of skilled managers, who achieve a positive alpha after management and transaction costs, diminishes rapidly during the study period. However, the proportion of unskilled managers increases during the same period. Using a refined false discovery approach, Ferson and Chen (2015) find smaller fractions of zero alpha managers and larger fractions of unskilled managers.

In summary, there is a contradiction between the growth of the actively managed mutual fund industry along with the discovery of some skill by active managers, and the empirical results of numerous studies that active management does not generate better performance than passive management. This contradiction leaves unresolved the question of whether active management adds value. A further issue raised by some studies, the benchmark choice problem, blurs this debate further.

4

1.2.3 Benchmark Choice Problem

As discussed previously, there are many approaches for measuring performance. Theoretically and empirically, choosing the right model for measurement, i.e., the benchmark choice problem, is an important issue. It is especially crucial in light of the empirical results that performance evaluation can change significantly across models and other methodological choices (Lehmann and Modest (1987) and Grinblatt and Titman (1994)).

In the context of the CAPM, Roll (1978), Dybvig and Ross (1985a, b) and Green (1986) show theoretically that performance measurement is sensitive to the benchmark choice. Choosing an inefficient benchmark may lead to biased performance evaluation. Lehmann and Modest (1987) confirm empirically this result by investigating the "bad model" problem for the CAPM and APT. Grinblatt and Titman (1994) also document the significant effect of the benchmark choice on Jensen's alpha and the measures of Grinblatt and Titman (1989) and Treynor and Mazuy (1966).

Chen and Knez (1996) argue that most beta-pricing models could fail to give zero performance to passive portfolios, a condition they establish for the admissibility of a performance measure. Ahn, Cao and Chrétien (2009) propose a conservative way to assess the effect of the benchmark choice problem. They examine performance values obtained by widely used models and show that between 8% and 50% of these values do not fall inside no-arbitrage bounds on admissible evaluation.

The literature interested in the benchmark choice problem shows that performance measures ultimately depend on the reliability of the underlying asset pricing model (see Fama (1998) for a discussion). One way to examine this reliability is by looking at the performance of performance measures in simulated experiments.

1.2.4 Comparison of Performance Measures with Simulations

There is a growing literature studying the validity of performance measures with simulations. Goetzmann, Ingersoll and Ivković (2000) show that the Henriksson and Merton (1981) timing measure applied to monthly returns of a simulated daily timer are

biased downward. Kothari and Warner (2001) argue that standard performance measures fail to detect abnormal performance, especially for mutual funds with style characteristics relatively different from the market portfolio. By constructing artificial mutual funds based on market timing or asset selectivity abilities, Farnsworth, Ferson, Jackson and Todd (2002) find little evidence of performance evaluation differences for SDF measures.

Using a bootstrap technique, Kosowski, Timmermann, Wermers and White (2006) find that some managers present selectivity skills and achieve positive performance after costs. They conclude that this performance persists and is not attributed to luck. Coles, Daniel and Nardari (2006) show that using a "bad model" leads to biased measures of selectivity and timing. The resulting invalid inferences may considerably affect fund ranking. Using a simulation setup inspired by Farnsworth, Ferson, Jackson and Todd (2002), Chrétien, Coggins and d'Amours (2015) study the performance of daily, occasional and monthly timers. They find that market timing measures are problematic when there is a mismatch between the frequency of timing activities and the frequency of data for performance measurement.

Relying on simulations with explicit controls on the true benchmark and manager's ability allows studying the effect of the benchmark choice problem and the power of performance measures to detect value added by active management. However, this literature focuses exclusively on standard models based on representative investors providing a unique mutual fund evaluation. Some recent studies argue instead on the importance of considering heterogeneous investors or clienteles whose different preferences can provide a multiplicity of mutual fund evaluations.

1.2.5 Heterogeneous Preferences and Investor Disagreement

Mutual fund investors select their investments based on numerous criteria, like management style, and clearly disagree on what funds are the most valuable for their portfolios. Accordingly, the vast majority of mutual funds advocate active management strategies to cater to specific clienteles of investors. However, most performance measures rely indirectly on representative investors by making assumptions on preferences, return distributions and/or market completeness. As argued by Ferson (2010), the real-life

6

existence of heterogeneous investor clienteles suggests the implementation of performance measures specific to the clienteles. The effect of investor heterogeneity and performance disagreement in mutual funds is an area of research with recent contributions by Ahn, Cao and Chrétien (2009), Bailey, Kumar and Ng (2011), Del Guercio and Reuter (2014) and Ferson and Lin (2014).

In a general setup where only the no arbitrage condition is imposed, Ahn, Cao and Chrétien (2009) find a set of admissible performance values that suggests large valuation disagreement across mutual funds investors. They argue that more than 80% of mutual funds are given a positive performance by some investors. Bailey, Kumar and Ng (2011) investigate reasons behind the variety of mutual funds across individual investors. They find that behavioral biases affect a certain class of investors, who trade frequently with bad timing and choose funds with relatively high expense ratios and turnover. This class of investors achieves poor performance as a result, compared to other classes of investors. They conclude that behavioral biases are a factor of investor heterogeneity in the mutual fund sector.

Del Guercio and Reuter (2014) distinguish between two types of market segments for retail funds: funds marketed directly to investors and funds sold through brokers. They find no evidence that funds marketed directly to retail investors underperform their benchmarks. However, they uncover evidence of underperformance of actively managed funds sold through brokers. They conclude that the retail mutual fund market is formed from two broad clienteles that value funds differently.

In an incomplete market setup, Ferson and Lin (2014) show that investor heterogeneity and disagreement affect considerably performance evaluation, as investors look differently at the attractiveness of the same fund. They develop bounds on expected disagreement with traditional alpha and document economically and statistically significant values. They argue that the effects of heterogeneous preferences and investor disagreement on performance are as important as the effects of the benchmark choice problem and the statistical imprecision in estimates of alpha.

As mutual fund investors have heterogeneous preferences, numerous investment styles are offered, with different styles reaching different clienteles. A literature exists on identifying and understanding mutual fund styles.

1.2.6 Mutual Fund Styles

Mutual funds cater to specific clienteles through their investment style (value, growth, large-cap, small-cap, etc.). Investors thus have a variety of styles to choose from, and style classifications, like the popular Morningstar Mutual Fund Style Box, are useful tools in their decision making process. Following Sharpe's (1992) introduction of return-based style analysis as a technique for uncovering the style allocations of mutual funds, there is a literature on the validity of existing mutual fund style classifications. Brown and Goetzmann (1997), Dibartolomeo and Witkowski (1997) and Kim, Shukla and Thomas (2000) show evidence of misclassification in reported mutual fund investment styles.

Mutual fund styles are also potentially helpful in identifying meaningful investor behavior and clienteles. Barberis and Shleifer (2003) argue that fund investors withdraw flows from styles having past poor performance, to invest in styles that exhibit high past performance. Blackburn, Goetzmann and Ukhov (2013) show that clientele characteristics are behind investors' preferences toward a specific style. They explain that investors in different styles exhibit significant differences in their trading behavior. Growth fund investors rely upon momentum, whereas value fund investors prefer contrarian trading.

Finally, there are biases in mutual fund data that can affect all empirical research on the topic. The last subsection of this review discusses the literature on the two most important biases.

1.2.7 Mutual Fund Data Biases

Although there have been important advances in historical data availability since Morningstar first provided electronic information on mutual funds, databases continue to suffer from biases affecting performance evaluation (Cuthbertson, Nitzsche and O'Sullivan (2010)). The first widely documented bias, as discussed by Brown, Goetzmann, Ibbotson and Ross (1992), Brown and Goetzmann (1995), Malkiel (1995), Gruber (1996) and

Carhart (1997), among others, is survivorship bias. Survivorship bias occurs when a poorly performing fund leaves the database so that researchers study only the performance and persistence of surviving funds (Aragon and Ferson (2006)). Brown, Goetzmann, Ibbotson and Ross (1992) show that survivorship bias affects the relationship between fund returns and their volatility. Malkiel (1995) documents the impact of survivorship bias, finding an average annual return of 15.69% for all funds versus 17.09% for surviving funds, between 1982 and 1991. To alleviate the survivorship bias documented in the Morningstar database, Brown and Goetzmann (1995) and Carhart (1997) construct samples which do not suffer from this bias, the latter's work leading to the creation of the now widely used CRSP Survivorship Bias Free Mutual Fund Database.

Another mutual fund database problem consists of including returns realized before the fund entry into the database, a back-fill bias discussed by Elton, Gruber and Blake (2001) and Aragon and Ferson (2006). Such returns can occur because of an incubation strategy where multiple funds are started privately, but only some funds are opened to the public at the end of the incubation period. Evans (2010) shows that an incubation bias affects the performance upward, as funds in incubation that eventually open to the public outperform already public funds by 3.5% annually during the incubation period. As a solution, Evans (2010) proposes to exclude the incubation period from the sample, a period that typically corresponds to the first year of available fund returns.

As reflected by the large number of contributions mentioned in this selective literature review, the evaluation of mutual fund performance is one of the most studied issues in finance. Despite the abundance of measures and results, this area of research is still very active and continues to evolve rapidly. We present in the next subsection the research questions, objectives and contributions to the existing literature of this thesis.

1.3   Research Questions, Objectives and Contributions

1.3.1   Research Questions

Mutual funds are among the preferred investment vehicles of investors. Of great importance to individuals, they often play a strategic role in their personal finances, either

through direct investment from their individual accounts or indirect use by pension funds. Furthermore, differentiated by their investment style and other management choices, they hope to satisfy the various needs of different clienteles. As previously stated, Ferson (2010, p.229) argues that it is important "to identify and characterize meaningful investor clienteles and to develop performance measures specific to the clienteles". In this thesis, we propose, develop and empirically implement a new performance evaluation approach with investor disagreement to answer this challenge. This approach is used to study three general research questions, one for each essay of this thesis. We can formulate these three questions as follows.

Essay 1 (Chapter 2): What is the performance evaluation of mutual funds from the perspective of their best potential clienteles?

Several empirical studies, including Glode (2011), Bailey, Kumar and Ng (2011) and Del Guercio and Reuter (2014), concentrate on identifying specific clienteles. Other papers, like Ferson and Lin (2014), study the general effect of investor disagreement and heterogeneity on mutual fund performance evaluation. However, these studies do not focus on the evaluation of mutual funds from the perspective of their best potential clienteles. Arguably, this evaluation is the most relevant performance evaluation for mutual funds. Although other investors may disagree with this evaluation, mutual funds should at least provide positive performance to their best clienteles. Using a measure based on an upper bound on admissible performance evaluation, extending Cochrane and Saá-Requejo (2000) and Ahn, Cao and Chrétien (2009), essay 1 develops and implements a measure of performance for the class of investors most favorable to mutual funds. It also provides additional evidence on investor disagreement and clientele effects in performance evaluation.

In addition to performance evaluation, we can use this new measure to develop a diagnostic tool for candidate performance measures and examine performance disagreement between two important groups of investors: *representative investors*, on whom standard measures typically rely for evaluation, versus *best clienteles*, who are potentially the most valuable targets for funds catering to specific clienteles.

Essay 2 (Chapter 3): How do best potential clienteles and representative investors implicit in commonly used asset pricing models differ in their performance evaluations of mutual funds?

Some studies use simulations with controlled managerial ability to diagnose the validity of performance measures (Kothari and Warner (2001), Farnswoth, Ferson, Jackson and Todd (2002), Kosowski, Timmermann, Wermers and White (2006) and Chrétien, Coggins and d'Amours (2015)). In contrast to these studies and aligned with the bound approach initiated by Hansen and Jagannathan (1991), essay 2 proposes a different diagnostic tool by using the best clientele alpha developed in essay 1. It then empirically implements the tool to assess the validity of the following candidate models: the CAPM and the models of Fama and French (1993), Carhart (1997) and Ferson and Schadt (1996)), four conditional linear factor models (conditional versions of the previous four models), two consumption-based models (a power utility model and an external habit-formation preference model first examined by Cochrane and Hansen (1992)), the manipulation proof performance measure of Goetzmann, Ingersoll, Spiegel and Welch (2007) and the nonparametric law-of-one-price (LOP) measure of Chen and Knez (1996).

Standard performance measures (like the CAPM alpha) focus on a unique benchmark (like the market portfolio) to represent the perspective of all investors. The best clientele alpha instead evaluates mutual funds from the perspective of their most favorable investors. Two alternative hypotheses can shed light on the difference between the alphas of representative investors and best clienteles. On the one hand, an inadmissibility problem, which is related to the benchmark choice problem, occurs when a candidate alpha is greater than the upper admissible bound that is the best clientele alpha. On the other hand, a misrepresentation problem occurs when a candidate alpha is lower than the best clientele alpha, and it indicates large investor disagreement in performance evaluation. As shown by Ferson and Lin (2014), with investor disagreement, the performance from standard measures can greatly misrepresent the value of funds for some of their clienteles. By formally comparing performance values for best potential clienteles with those for representative investors, essays 2 quantifies this misrepresentation and compares it across

different standard measures. Such a comparison forms the basis of an evaluation of the appropriateness of standard measures for the purpose of performance evaluation.

Another application of the best clientele performance approach developed in essay 1 is to focus on clienteles interested in particular investment styles within the mutual fund industry for performance measurement.

Essay 3 (Chapter 4): What is the performance evaluation of mutual funds from the perspective of clienteles with favorable preferences for different investment styles?

Equity style investing (value, growth, small-cap, large-cap) has become dominant in industry practices. For example, thousands of equity mutual funds advertised themselves according to their size and value focuses, oftentimes starting with their names. They cater to and attract millions of size and value investors. Equity styles are thus relevant to identify meaningful investor clienteles. Barberis and Shleifer (2003) point out the growing interest of financial service firms to understand style preferences and there is a related literature on existing mutual fund style classifications (see Kim, Shukla, and Tomas (2000), Brown and Goetzmann (1997), Dibartolomeo and Witkowski (1997), among others). Some studies attempt to link the decision to invest in particular styles to investors' preferences (see Del Guercio and Tkac (2002) and Goetzmann and Massa (2002)). Other studies, like Shefrin and Statman (1995, 2003), Blackburn, Goetzmann and Ukhov (2013) and Shefrin (2015) document significant differences in the judgments, sentiment sensitivity and trading behavior of style investors.

Essay 3 develops and implements style-clientele-specific performance measures to account for the style preferences of mutual fund investors. The approach is based on a new method to better exploit existing style classification data and group funds into representative style portfolios that identify meaningful style clienteles. Then, we explore the economic properties of the marginal style preferences extracted from our approach. Finally, we use these implied preferences in a clientele-specific performance evaluation of individual mutual funds.

1.3.2 Objectives and Contributions

The aim of this thesis is to present three essays allowing:

- To improve the performance evaluation of mutual funds by accounting for investor disagreement and focusing on the best potential clienteles. Using asset pricing bounds, we propose, develop and implement a measure of mutual fund performance from the point of view of the most favorable classes of investors. The results contribute to the previously reviewed literatures on performance evaluation approaches, the value added by active management, heterogeneous preferences and investor disagreement in mutual funds. They provide an answer to Ferson's (2010) challenge to implement a performance measure specific to clienteles.

- To formally compare the evaluations from the best clientele performance measure with those from standard performance measures implicitly based on representative investors. Using this comparison as a diagnostic tool, we document whether the inadmissibility and misrepresentation problems are important issues for candidate performance measures. The results contribute to the previously reviewed literatures on the benchmark choice problem, the comparison of performance measures with simulations and heterogeneous preferences and investor disagreement in mutual funds. In particular, they give an examination of the validity of standard measures using an approach different from the simulation-based methods proposed in the literature. The results also contribute to the manipulation proof performance measure literature because we provide a new estimation strategy that allows statistical inferences on the significance of the manipulation proof performance values.

- To improve the performance evaluation of mutual funds by implicitly considering the preferences of clienteles favorable to different equity investment styles. We implement clientele-specific performance measures and explore how different are the preferences and evaluations of investors attracted to various fund styles. The results contribute to the previously reviewed literatures on performance evaluation approaches, the benchmark choice problem, heterogeneous preferences and investor disagreement, and mutual fund styles. In particular, we provide supporting evidence

13

for the conjecture of Ferson (2010) that clientele-specific measures based on meaningful investor clienteles might be necessary to properly evaluate mutual funds. We also contribute to the mutual fund style literature by developing a new style classification method to better exploit existing objective code data and account for code changes and missing codes.

1.4 The Best Clientele Performance Evaluation Approach

The performance approach developed in this thesis starts by measuring the performance, or alpha, with the stochastic discount factor (SDF) approach such that:

$$\alpha_{MF} = E[m\,R_{MF}] - 1,$$

(1)

where m is the SDF of an investor interested in valuing the mutual fund with return R_{MF}. Glosten and Jagannathan (1994) and Chen and Knez (1996) are the first to propose the SDF alpha for performance evaluation. Ferson (2010) summarizes its benefits.

We then impose an economically relevant structure on the set of SDFs of all investors to obtain a restricted set useful to identify the most favorable performance. Let M represents this restricted set. Under the assumption that this set is constrained enough to be closed and convex, Chen and Knez (1996) and Ahn, Chrétien and Cao (2009) demonstrate that it is possible to find an upper bound on the performance of a fund:

$$\bar{\alpha}_{MF} = \sup_{m \in M} E[m\,R_{MF}] - 1,$$

(2)

where $\bar{\alpha}_{MF}$ represents the upper bound on the alpha, i.e., the highest average performance value that can be found from the heterogeneous investors considered.

By considering a set of SDFs, as opposed to selecting a unique SDF, our setup results in a finite range of performance values (see Chen and Knez (1996) and Ahn, Chrétien and Cao (2009)). Hence, it allows for investor disagreement that can occur in an incomplete market (Chen and Knez (1996) and Ferson and Lin (2014)). One key to our

14

approach is restricting the set M of SDFs in an economically meaningful way. We impose two insightful conditions that the SDFs of mutual fund investors should meet: the law-of-one-price and no-good-deal conditions.

The law-of-one-price condition is discussed extensively by Hansen and Jagannathan (1991) and states that the SDFs used for performance measurement should correctly price passive portfolios or basis assets:

(3)
$$E[m\,\mathbf{R_K}] - \mathbf{1} = 0,$$

where $\mathbf{R_K}$ is a vector of returns on K passive portfolios, and $\mathbf{1}$ is a $K \times 1$ unit vector. It is plausible to assume that mutual fund investors would agree that passively-managed portfolios should have zero average alphas. The main benefit of imposing the law-of-one-price condition is to alleviate the previously mentioned benchmark choice problem.

Although the law-of-one-price condition provides important restrictions on the set of SDFs, it is not sufficient to make it closed and convex, and thus would allow an infinite range of performance values, as discussed in Chen and Knez (1996). The second condition resolves this issue. The no-good-deal condition of Cochrane and Saá-Requejo (2000) states that the SDFs used for performance measurement should not allow investment opportunities with Sharpe ratios that are too high:

(4)
$$\frac{E[R_j - R_F]}{\sigma[R_j - R_F]} < \overline{h},$$

where R_j is the return on any potential asset j, $\overline{h}$ is the maximum Sharpe ratio allowable and R_F is the risk-free rate. We thus stipulate that mutual fund investors would find it implausible that investment opportunities could provide Sharpe ratios that are too high, making them good deals (see Ross (1976), MacKinlay (1995), Cochrane and Saá-Requejo (2000) and Ross (2005)).

By extension of the analysis of Hansen and Jagannathan (1991), Cochrane and Saá-Requejo (2000) demonstrate that the law-of-one-price and no-good-deal conditions restrict the set M of SDFs for mutual fund investors by limiting their variability:

$$(5) \qquad \frac{h^*}{R_F} \leq \sigma(m) \leq \frac{\bar{h}}{R_F},$$

where h^* is the optimal Sharpe ratio attainable from the passive portfolios. Intuitively, we can interpret the SDFs as representing the marginal preferences of mutual fund investors. This restriction stipulates that the variability of managerial utilities should be large enough to correctly price existing passive portfolios, but small enough to rule out implausibly high risk aversion that would allow good deals be viable.

Cochrane and Saá-Requejo (2000) also show that this restriction makes the set of SDFs closed and convex, and thus allows for finite price bounds. The approach developed in this thesis adopts their framework to performance measurement. Specifically, we obtain an upper bound on performance evaluation that we call the "best clientele performance evaluation or alpha". The best clientele alpha possesses an analytical solution that has $2K + 2$ parameters to estimate, where K is the number of passive portfolios used to impose the law-of-one-price condition. The closed-form solution can be estimated with the generalized method of moments of Hansen (1982).

1.5   Conclusion

There is a large and still growing literature on mutual fund performance evaluation. In this chapter, we introduce the general issues motivating this thesis, present its main research questions, objectives and contributions, and provide an overview of the best clientele performance approach it develops.

The rest of the thesis is divided as follow. In chapter 2, essay 1 provides the first comprehensive performance evaluation exercise for best potential clienteles of mutual funds. In chapter 3, essay 2 focuses on a new diagnostic tool for candidate performance models and its associated inadmissibility and misrepresentation problems. In chapter 4,

essay 3 develops clientele-specific performance measures based on the style preferences of mutual fund investors. The last chapter offers concluding remarks.

1.6 References

Ahn, D., Cao, H. H., Chrétien, S., (2009), "Portfolio performance measurement: a no arbitrage bounds approach", *European Financial Management*, 15, 298-339.

Aragon, G., Ferson, W., (2006), "Portfolio performance evaluation", *Foundations and Trends in Finance*, 2, 101-111.

Bailey, W., Kumar, A., Ng, D., (2011), "Behavioral biases and mutual fund investors", *Journal of Financial Economics*, 102, 1-27.

Baks, K. P., Metrick, A., Wachter, J., (2001), "Should investors avoid all actively managed mutual fund? A study in Bayesian performance evaluation", *Journal of Finance*, 56, 45-85.

Barberis, N., Shleifer, A., (2003), "Style investing", *Journal of Financial Economics*, 68, 161-199.

Barras, L., Scaillet, O., Wermers, R., (2010), "False discoveries in mutual fund performance: Measuring luck in estimated alphas", *Journal of Finance*, 65, 179-216.

Berk, J. B., (2005), "Five myths of active portfolio management", *Journal of Portfolio Management*, 31, 27-31.

Berk, J. B., Green, R. C., (2004), "Mutual fund flows and performance in rational markets", *Journal of Political Economy*, 112, 1269-1295.

Blackburn, D. W., Goetzmann, W. N., Ukhov, A. D., (2014), "Is trading behavior stable across contexts? Evidence from style and multi-style investors", *Quantitative Finance*, 14, 605-627.

Brown, J. S., Goetzmann, W. N., (1995), "Performance persistence", *Journal of Finance*, 50, 679-698.

Brown, S. J., Goetzmann, W. N., (1997), "Mutual fund styles", *Journal of Financial Economics*, 43, 373-399.

Brown, S. J., Goetzmann, W. N., Ibbotson, R. G., Ross, S. A., (1992), "Survivorship bias in performance studies", *Review of Financial Studies*, 5, 553-580.

Carhart, M. M., (1997), "On persistence in mutual fund performance", *Journal of Finance*, 52, 57-82.

Chen, Z., Knez, P. J., (1996), "Portfolio performance measurement: theory and applications", *Review of Financial Studies*, 9, 511-555.

Chrétien, S., Coggins, F., d'Amours, F., (2015), "The performance of market timing measures in a simulated environment", Unpublished Working Paper, Laval University.

18

Christopherson, J. A., Ferson, W. E., Glassman, D. A., (1998), "Conditioning manager alphas on economic information: Another look at the persistence of performance", *Review of Financial Studies*, 11, 111-142.

Cochrane, J., Hansen, L. P., (1992), "Asset pricing explorations for macroeconomics", *NBER Macoeconomics Annual*, 7, 115-165.

Cochrane, J., Saá-Requejo, J., (2000), "Beyond arbitrage: good-deal asset price bounds in incomplete markets", *Journal of Political Economy*, 108, 79-119.

Cogneau, P., Hübner, G., (2009a), "The (more than) 100 ways to measure portfolio performance part 1: Standardized risk-adjusted measures", *Journal of Performance Measurement*, 13, 56-71.

Cogneau, P., Hübner, G., (2009b), "The (more than) 100 ways to measure portfolio performance part 2: Special measures and comparison", *Journal of Performance Measurement*, 14, 56-69.

Coles, J. L., Daniel, N. D., Nardari, F., (2006), "Does the misspecification of timing strategy or timing index affect inference in measuring mutual fund performance?", Unpublished Working Paper, Arizona State University.

Cremers, K. J. M., Petajisto, A., (2009), "How active is your fund manager? A new measure that predicts performance », *Review of Financial Studies*, 22, 3329-3365.

Cuthbertson, K., Nitzsche, D., O'Sullivan, N., (2010), "Mutual fund performance: measurement and evidence", *Journal of Financial Markets*, *Institutions and Instruments*, 19, 95-187.

Dahlquist, M., Söderlind, P., (1999), "Evaluating portfolio performance with stochastic discount factors", *Journal of Business*, 72, 347-383.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., (1997), "Measuring mutual fund performance with characteristic-based benchmarks", *Journal of Finance*, 52, 1035-1058.

Del Guercio, D., Reuter, J., (2014), "Mutual fund performance and the incentive to generate alpha ", *Journal of Finance,* 69, 1673-1704.

Del Guercio, D., Tkac, P. A., (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.

Dibartolomeo, D., Witkowski, E., (1997), "Mutual fund misclassification: evidence based on style analysis", *Financial Analysts Journal*, 53, 32-43.

Dybvig, P. H., Ross, S. A., (1985a), "The analytics of performance measurement using a security market line", *Journal of Finance*, 40, 401-416.

Dybvig, P. H., Ross, S. A., (1985b), "Differential information and performance measurement using a security market line", *Journal of Finance*, 40, 401-416.

Elton, E. J., Gruber, M. J., Blake, C. R. (2001), "A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and morningstar mutual fund databases", *Journal of Finance*, 56, 2415-2430.

Elton, E. J., Gruber, M. J., Das, S., Hlavka, M., (1993), "Efficiency with costly information: A reinterpretation of evidence from managed portfolios", *Review of Financial Studies*, 6, 1-22.

Evans, R. B., (2010), "Mutual fund incubation", *Journal of Finance*, 65, 1581-1611.

Fama, E. F., (1998), "Market efficiency, long-term returns and behavioral finance", *Journal of Financial Economics*, 49, 283-306.

Fama, E. F., French, K. R., (1993), "Common risk factors in the returns on stocks and bouds", *Journal of Financial Economics*, 33, 3-56.

Fama, E. F., French, K. R., (2010), "Luck versus skill in the cross-section of mutual fund returns", *Journal of Finance*, 65, 1915-1948.

Farnsworth, H., Ferson, W., Jackson, D., Todd, S., (2002), "Performance evaluation with stochastic discount factors", *Journal of Business*, 75, 473-503.

Ferson, W. E., (2010), "Investment performance evaluation", *Annual Review of Financial Economics*, 2, 207-234.

Ferson, W. E., Chen, Y., (2015) "How many good and bad fund managers are there, really?" Unpublished Working Paper, University of Southern California and Texas A&M University.

Ferson, W. E., Lin, J., (2014), "Alpha and performance measurement: the effect of investor heterogeneity", *Journal of Finance*, 69, 1565-1596.

Ferson, W. E., Schadt, R., (1996), "Measuring fund strategy and performance in changing economic conditions", *Journal of Finance*, 51, 425-462.

Ferson, W., Henry, T., Kisgen, D., (2006), "Evaluating government bond fund performance with stochastic discount factors", *Review of Financial Studies*, 19, 423-455.

Ferson, W., Khang, K., (2002), "Conditional performance measurement using portfolio weights: Evidence for pension funds", *Journal of Financial Economics*, 65, 249-82.

French, K. R., (2008), "Presidential address: The cost of active investing", *Journal of Finance*, 63, 1537-1573.

Glode, V., (2011), "Why mutual funds underperform", *Journal of Financial Economics*, 93, 546-559.

20

Glosten, L., Jagannathan, R., (1994), "A contingent claims approach to performance evaluation", *Journal of Empirical Finance*, 1, 133-66.

Goetzmann, W. N., Ingersoll Jr., J., Ivković, Z., (2000), "Monthly measurement of daily timers", *Journal of Financial and Quantitative Analysis*, 35, 257-290.

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

Goetzmann, W. N., Massa, M., (2002), "Daily momentum and contrarian behavior of index funds investors", *Journal of Financial and Quantitative Analysis*, 37, 375-389.

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

Grinblatt, M., Titman, S., (1989), "Mutual fund performance: An analysis of quarterly portfolio holdings", *Journal of Business*, 62, 393-416.

Grinblatt, M., Titman, S., (1993), "Performance measurement without benchmarks: An examination of mutual fund returns", *Journal of Business*, 66, 47-68.

Grinblatt, M., Titman, S., (1994), "A study of monthly mutual fund returns and performance evaluation techniques", *Journal of Financial and Quantitative Analysis*, 29, 419-444.

Gruber, M. J., (1996), "Another puzzle: The growth in actively managed mutual funds", *Journal of Finance*, 51, 783-810.

Hansen, L. P., (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029-1054.

Hansen, L. P., Jagannathan, R., (1991), "Implication of security market data for models of dynamic economies", *Journal of Political Economy*, 99, 225-262.

Henriksson, R. D., Merton, R. C., (1981), "On market timing and investment performance: Statistical procedures for evaluating forecasting skills", *Journal of Business*, 54, 513-533.

Jensen, M. C., (1968), "The performance of mutual funds in the period 1945-1964", *Journal of Finance*, 23, 389-416.

Kacperczyk, M., Sialm, C., Zheng, L., (2005), "On the industry concentration of actively managed equity mutual funds", *Journal of Finance*, 60, 1983-2011.

Kim, M., Shukla, R., Tomas, M., (2000), "Mutual fund objective misclassification", *Journal of Economics and Business*, 52, 309-323.

Kosowski, R., Timmermann, A., Wermers, R., White, H. L., (2006), "Can mutual fund "stars" really pick stocks? Evidence from a bootstrap analysis", *Journal of Finance*, 61, 2251-2295.

Kothari, S. P., Warner, J. B., (2001), "Evaluating mutual fund performance", *Journal of Finance*, 56, 1985-2010.

Lehmann, B. N., Modest, D. M., (1987), "Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons", *Journal of Finance*, 42, 233-66.

Mackinlay, A. C., (1995), "Multifactor models do not explain deviations from the CAPM", *Journal of Financial Economics*, 38, 3-28.

Malkiel, B. G., (1995), "Returns from investing in equity mutual funds 1971 to 1991", *Journal of Finance*, 50, 549-572.

Pástor, L., Stambaugh, R. F., (2002a), "Mutual fund performance and seemingly unrelated assets", *Journal of Financial Economics*, 63, 315–349.

Pástor, L., Stambaugh, R. F., (2002b), "Investing in equity mutual funds", *Journal of Financial Economics*, 63, 351–380.

Roll, R., (1978), "Ambiguity when performance is measured by the security market line", *Journal of Finance*, 33, 1051-1069.

Ross, S. A., (1976), "The arbitrage theory of capital asset pricing", *Journal of Economics Theory*, 13, 341-360.

Ross, S. A., (2005), "Neoclassical finance", Princeton University Press, Princeton, NJ.

Sharpe, W. F., (1966), "Mutual fund performance", *Journal of Business*, 1, 119-138.

Sharpe, W. F., (1992), "Asset allocation: management style and performance measurement", *Journal of Portfolio Management*, 18, 7-19.

Shefrin, H., (2015), "Investors' judgments, asset pricing factors and sentiment", *European Financial Management*, 2, 205-227.

Shefrin, H., Statman, M., (1995), "Making sense of beta, size and book-to-market", *Journal of Portfolio Management*, 21, 26-34.

Shefrin, H., Statman, M., (2003), "The style of investor expectations", in Coggin, T., Fabozzi, F. (eds), *Handbook of Equity Style Management*, Wiley, 195-218.

Treynor, J., (1965), "How to rate management of investment funds", *Harvard Business Review*, 43, 63-75.

Treynor, J., Mazuy, K., (1966), "Can mutual funds outguess the market?", *Harvard Business Review*, 44, 131-136.

2   Mutual Fund Performance Evaluation and Best Clienteles

Abstract

This paper investigates investor disagreement and clientele effects in performance evaluation by developing a measure that considers the best potential clienteles of mutual funds. In an incomplete market under law-of-one-price and no-good-deal conditions, we obtain an upper bound on admissible performance measures that identifies the most favorable alpha. Empirically, we find that a reasonable investor disagreement leads to generally positive performance for the best clienteles. Performance disagreement by investors can be significant enough to change the average evaluation of mutual funds from negative to positive, depending on the clienteles.

Keywords: Portfolio Performance Measurement; Clienteles; Investor Disagreement; No-Good-Deal Bounds

Résumé

Cet article étudie le désaccord entre investisseurs et les effets de clientèle dans l'évaluation de performance en développant une mesure qui considère les meilleures clientèles potentielles des fonds mutuels. Dans un marché incomplet sous conditions de la loi d'un seul prix et d'absence de bonnes affaires, nous obtenons une borne supérieure sur les mesures de performance admissibles qui identifie l'alpha le plus favorable. Empiriquement, nous trouvons qu'un désaccord raisonnable entre investisseurs résulte en une performance généralement positive pour les meilleures clientèles. Le désaccord entre investisseurs peut être suffisamment significatif pour changer l'évaluation moyenne des fonds mutuels de négative à positive, dépendamment des clientèles.

2.1   Introduction

In today's mutual fund industry, there are thousands of funds that cater to different investors through their management style and other attributes. In incomplete markets, as investors can disagree about the attractiveness of funds, this catering might be worthwhile in leading the funds to find appropriate clienteles, i.e., the class of investors to whom they are the most valuable.

Recent research examines the effect of investor disagreement and heterogeneity on mutual fund performance evaluation. Studying the issue generally, Ferson and Lin (2014) find that taking into account heterogeneous preferences can lead to large valuation disagreement. In particular, they develop a bound on expected disagreement with a traditional alpha and show that such disagreement can be similar in importance to the widely documented effects of the benchmark choice problem and the statistical imprecision in estimates of alpha. They furthermore provide evidence that investor disagreement and heterogeneity are economically significant in the behavior of fund investors.

Many studies concentrate more specifically on identifying specific clienteles. Glode (2011) argues that mutual funds could be valuable to investors with high marginal utilities in difficult times by providing positive alphas in recessions. Bailey, Kumar and Ng (2011) document that behavioral biases are factors of investor heterogeneity in the mutual fund industry. Del Guercio and Reuter (2014) find that the retail mutual fund market is formed from two broad clienteles that value funds differently: self-directed investors and investors acting with the help of brokers. In a literature review, Ferson (2010) emphasizes that measuring performance from the point of view of different clienteles is a challenge for future research.

Despite these contributions, the literature has not focused on the valuation that can be the most important for mutual funds, the one from their best potential clienteles. The goal of this paper is to provide additional evidence on investor disagreement and clientele effects in performance evaluation by developing and implementing a measure that considers the best potential clienteles of mutual funds. For our purpose, these clienteles are defined as the class of investors most favorable to a fund in the sense that they value the

fund at an upper performance bound in a setup where the market is incomplete. Our "best clientele performance measure" thus not only considers investor disagreement but also focuses on the most worthy clienteles that a mutual fund could target.

We develop this new measure by combining the asset pricing bound literature with the stochastic discount factor (SDF) performance evaluation approach first proposed by Glosten and Jagannathan (1994) and Chen and Knez (1996). Specifically, Cochrane and Saá-Requejo (2000) propose asset pricing bounds in an incomplete market under the law-of-one-price condition of Hansen and Jagannathan (1991) and a no-good-deal condition that rules out investment opportunities with unreasonably high Sharpe ratios, termed "good deals". We obtain the best clientele alpha by adapting this approach to portfolio performance measurement and focusing on the upper performance bound.

Using Hansen's generalized method of moments (1982), we estimate the best clientele SDF alphas with monthly returns of 2786 actively managed U.S. open-ended equity mutual funds from January 1984 to December 2012. Our main results rely on a set of passive portfolios based on ten industry portfolios, although they are similar for style portfolios or the market portfolio. Following the literature, they consider the disagreement generated by allowing a maximum Sharpe ratio (specifying the no-good-deal restriction) equal to the best Sharpe ratio from the passive portfolios plus either half the Sharpe ratio of the market index or its full value.

Empirically, we find that considering investor disagreement and focusing on the best potential clienteles lead to a generally positive performance for mutual funds. For example, with a disagreement that corresponds to an increase in admissible opportunities equivalent to half the market Sharpe ratio, the mean monthly best clientele alpha is equal to 0.236% (*t*-stat. = 3.35). Comparatively, the mean alpha when disagreement is ruled out is -0.179% (*t*-stat. = -3.14), a value similar to the findings from standard measures based on representative investors. The spread of 0.415% between these values is comparable to the magnitude of the bound on the average disagreement documented by Ferson and Lin (2014). Accordingly, the proportions of positive and significantly positive alpha estimates increase from 20% to 78% and from 1% to 24%, respectively, when allowing for disagreement. To account for false discoveries, we implement the technique of Barras,

26

Scaillet and Wermers (2010) and its extension proposed by Ferson and Chen (2015). For the best clienteles, we find that the proportions of skilled funds increase considerably, and the proportions of unskilled funds disappear.

Further augmenting the maximum Sharpe ratio improves even more the performance of the mutual funds for their best clienteles, and the results are robust to the use of simulated finite sample distributions for inference purposes. In addition, we show that increasing monthly Sharpe ratio opportunities by only 0.04 (approximately one third of the market Sharpe ratio) is sufficient for the best clienteles to give a nonnegative performance to mutual funds on average. Hence, although the maximum Sharpe ratio is a subjective choice, our findings indicate that a reasonably small disagreement among investors is enough to generate a positive performance from the best potential clienteles of a majority of funds.

We also implement a conditional version of the best clientele measure and find that the inclusion of conditioning information does not alter our findings. Consistent with Moskowitz (2000), Kosowski (2011) and Glode (2011), we find evidence that the best clientele alpha estimates are more positive in recessions than in expansions. Finally, to further explore the issues of investor disagreement and clientele effects, we estimate a lower performance bound to obtain a "worst clientele alpha" and investigate the total performance disagreement resulting from the difference between best and worst clientele alphas. We find that the evidence from the worst clienteles is more negative than the one from standard measures and that total disagreement is economically and statistically significant.

Overall, the positive best clientele alpha estimates documented in this paper have two important implications for the mutual fund literature. First, as shown by Chen and Knez (1996) and Ferson and Lin (2014), a positive SDF alpha is required for the existence of investors who would want to buy a fund. However, a large number of studies document negative value added for actively managed mutual funds (see Fama and French (2010) and Barras, Scaillet and Wermers (2010) for recent examples), with the growth of the industry termed a "puzzle" by Gruber (1996). In line with the continued popularity of these funds in

practice, this paper empirically shows that a positive alpha exists for some clienteles for most funds.

Second, although the effects of the benchmark choice problem and the statistical imprecision in estimates of alpha can change the performance assessment of a fund, these issues do not generally change the conclusion regarding the negative value added for the mutual fund industry. This paper shows that investor disagreement can be significant enough to change the average evaluation of mutual funds from negative to positive, depending on the clienteles. It thus reinforces the analysis of Ferson and Lin (2014) on the economic importance of investor disagreement and clientele effects.

The remainder of this paper is organized as follows. Section 2.2 develops the best clientele performance measure. Section 2.3 presents our methodology for estimating performance values and summarizing the results. Section 2.4 describes the mutual fund data and the passive portfolio returns. Section 2.5 presents and interprets our empirical results. Finally, section 2.6 concludes.

2.2 Performance Measure for the Best Potential Clienteles

2.2.1 Basic Performance Setup

Our approach starts by measuring the performance, or alpha, with the stochastic discount factor (SDF) approach such that:

(1) $$\alpha_{MF,t} = E_t\big[m_{t+1}R_{MF,t+1}\big] - 1,$$

where m_{t+1} is the SDF of an investor interested in valuing the mutual fund with gross return $R_{MF,t+1}$, and the expectation operator $E_t[\ \ ]$ is understood to be conditional on the investor's or the public information set. Taking the unconditional expectation on both sides, and dropping time subscripts except when needed to avoid ambiguity, the expected alpha is given by:

28

$$(2) \qquad\qquad\qquad \alpha_{MF} = E[m\, R_{MF}] - 1.$$

Glosten and Jagannathan (1994) and Chen and Knez (1996) were the first to propose SDF alphas for performance evaluation. The SDF approach does not require any assumptions about complete markets, utility functions or aggregation. In the context of performance evaluation, Ferson (2010) argues that it is general enough to properly account for heterogeneous investors and differentially informed managers. In contrast, for a traditional regression-based alpha, a positive (negative) value does not necessarily imply buying (selling) the fund, and a manager with superior information does not necessarily generate a positive value. Ferson (2010, p. 227) thus concludes that « the SDF alpha seems to be on the most solid theoretical footing, and should probably get more attention than it has in the literature ».

Unfortunately, the SDFs of different investors are not observable. Hence, the literature typically uses the SDF from a representative investor obtained through economic assumptions and equilibrium conditions. This approach provides a unique performance evaluation that can be relevant for all investors. However, it rules out investor disagreement that occurs when one client views the performance of a fund differently from another client (see Ferson and Lin (2014)). Furthermore, it exposes the results to the benchmark choice problem because the selected performance model does not necessarily price correctly passive portfolios (see Chen and Knez (1996), Ahn, Cao and Chrétien (2009) and Cremers, Petajisto and Zitzewitz (2013)).

Instead, in this paper, we impose an economically relevant structure on the set of SDFs of all investors to obtain a restricted set useful to identifying the most favorable performance. Let M represent this restricted set. Under the assumption that it is constrained enough to be closed and convex, Chen and Knez (1996) and Ahn, Chrétien and Cao (2009) demonstrate that it is possible to find an upper bound on the performance of a fund:

$$(3) \qquad\qquad\qquad \bar{\alpha}_{MF} = \sup_{m \in M} E[m\, R_{MF}] - 1$$

where $\bar{\alpha}_{MF}$ represents the upper bound on the expected alpha, the highest average performance value that can be found from the heterogeneous investors considered in M.

By considering a set of SDFs, as opposed to selecting a unique SDF, our setup results in a finite range of performance values (see Chen and Knez (1996) and Ahn, Chrétien and Cao (2009)). Hence, it allows for investor disagreement that can occur in incomplete market. As argued by Chen and Knez (1996, p. 529), "given that mutual funds are set up to satisfy different clienteles, such an evaluation outcome may not be unrealistic". Empirically, the results of Ferson and Lin (2014) suggest that investor disagreement is in fact economically important. Although we also provide some results for the lower bound, we focus on the upper bound because it can be interpreted as the performance from the class of investors most favorable to the mutual fund (in a valuation sense). It is thus possible to evaluate whether mutual funds add value from the perspective of their best potential clienteles. In particular, Chen and Knez (1996) and Ferson and Lin (2014) show that if this value is positive, there are some investors who would want to buy the fund, with an optimal investment proportional to the alpha.

2.2.2   Restricting the Stochastic Discount Factors

One key to our approach is restricting the set of SDFs in an economically meaningful way. We impose two conditions that the SDFs of mutual fund investors should meet. The first condition is the law-of-one-price condition, as discussed by Hansen and Jagannathan (1991): The SDFs used for performance measurement should price correctly passive portfolios or basis assets:

$$(4) \qquad\qquad E[m\,\mathbf{R_K}] - \mathbf{1} = 0,$$

where $\mathbf{R_K}$ is a vector of gross returns on K passive portfolios, and $\mathbf{1}$ is a $K \times 1$ unit vector. It is plausible to assume that mutual fund investors would agree that passively managed portfolios should have zero average alphas. The main benefit of imposing the law-of-one-price condition is to alleviate the previously mentioned benchmark choice problem. To see how this condition restricts the set of SDFs for investors, Hansen and Jagannathan (1991)

30

provide the best known bound by showing that it implies a minimum SDF standard deviation that is related to the highest Sharpe ratio attainable in the passive portfolios.[1]

Although the law of one price provides important restrictions on the set of SDFs, it is not sufficient to make it closed and convex and thus would allow an infinite range of performance values, as discussed in Chen and Knez (1996). The second condition we impose is the no-good-deal condition of Cochrane and Saá-Requejo (2000). Specifically, the SDFs should not allow investment opportunities with Sharpe ratios that are too high:

$$\text{(5)} \qquad\qquad \frac{E[R_j - R_F]}{\sigma[R_j - R_F]} < \bar{h},$$

where R_j is the return on any potential asset j, $\bar{h}$ is the maximum Sharpe ratio allowable, and R_F is the risk-free rate. We thus stipulate that investors would find it implausible that investment opportunities could provide Sharpe ratios that are too high, making them good deals. If available, these deals would be unlikely to survive because investors would quickly grab them up.

The reasons why high Sharpe ratios should be ruled out are discussed by Ross (1976), MacKinlay (1995), Cochrane and Saá-Requejo (2000) and Ross (2005), among others. Ross (1976) argues that Sharpe ratios that are too high (more than twice the market Sharpe ratio) are unreasonable from the perspective of the CAPM and thus rules them out in studying deviations from the arbitrage pricing theory. In developing a specification test for multifactor models, MacKinlay (1995) uses a bound on the maximum Sharpe ratio, arguing that high ratios are unlikely from the perspective of risk-based models. Cochrane and Saá-Requejo (2000) argue that implausibly high Sharpe ratio opportunities should be rapidly exploited by investors. Unless there are limits to exploiting them, their presence would imply implausibly high investor risk aversion. A similar argument is formalized in Ross (2005).

---

[1]Other restrictions on SDFs that can be developed from this condition include, for examples, the bounds of Snow (1991) on selected higher SDF moments and the bound of Chrétien (2012) on the SDF autocorrelation.

By extension of the analysis of Hansen and Jagannathan (1991), Cochrane and Saá-Requejo (2000) demonstrate that the no-good-deal condition restricts the set of SDFs for mutual fund investors by limiting its second moment:

$$
(6) \qquad E[m^2] \leq \frac{\left(1 + \bar{h}^2\right)}{R_F{}^2}.
$$

Put differently, the assumption that investors would necessarily want to take part at the margin in implausibly good deals implies that their SDF satisfies the restriction $\sigma[m] \leq \bar{h}/R_F$. As shown by Cochrane and Saá-Requejo (2000), this restriction and the law-of-one-price condition make the set of SDFs closed and convex and thus allow the existence of price bounds.

There are other restrictions that could be imposed on SDFs to result in price bounds.[2] In this paper, we select the no-good-deal condition for the following reasons. First, since being introduced by Sharpe (1966), the Sharpe ratio has a long history of relevancy in performance evaluation. Due to its simplicity and intuitive appeal, it is a commonly used measure both in practice and in academic studies. Second, the literature offers some guidance on the choice of the maximum Sharpe ratio. In contrast, there is little guidance for the maximum gain-loss ratio, and there are often not enough restrictions imposed by the no-arbitrage condition (so no-arbitrage bounds are typically wide). Third, the Sharpe ratio captures approximate arbitrage opportunities as well as the gain-loss ratio when returns are normally distributed, which is reasonable for our sample of monthly equity mutual fund returns.[3] Fourth, the no-good-deal framework offers a closed-form

---

[2]For example, Hansen and Jagannathan (1991) discuss the no arbitrage condition that excludes non-positive SDFs by ruling out arbitrage opportunities. They show that it restricts further the set of SDFs by increasing the SDF volatility bound. Chen and Knez (1996) show that it is sufficient to obtain a finite range of performance values, and Ahn, Cao and Chrétien (2009) study no arbitrage performance bounds for mutual funds. As an alternative, Bernardo and Ledoit (2000) introduce a maximum gain-loss ratio condition that rules out approximate arbitrage opportunities. They show that the condition leads to a restriction on the minimum and maximum SDF values.

[3]See, for example, the results of Koski and Pontiff (1999) that most equity mutual funds do not use derivatives and that the distributional characteristics of users and non-users are similar, with monthly skewness and excess kurtosis that are relatively close to zero.

solution for the performance bounds. This solution facilitates and accelerates its implementation in comparison to numerical-only solutions obtained when imposing the no arbitrage condition or the maximum gain-loss ratio condition. This advantage should not be neglected given that our large-scale empirical investigation considers thousands of mutual funds.[4]

2.2.3 Best Clientele Performance Measure

Considering our basic performance setup and our restrictions on the set of SDFs, the upper bound on performance evaluation can be found by solving the following problem:

$$\bar{\alpha}_{MF} = \sup_{m \in M} E[m \, R_{MF}] - 1, \tag{7}$$

$$\text{subject to } E[m \, \mathbf{R_K}] = \mathbf{1}, \; E[m^2] \leq \frac{(1+\bar{h}^2)}{R_F{}^2}. \tag{8}$$

Cochrane and Saá-Requejo (2000) show that this problem has the following solution:

$$\bar{\alpha}_{MF} = E[\bar{m} R_{MF}] - 1, \tag{9}$$

with:

$$\bar{m} = m^* + vw \tag{10}$$

$$m^* = \mathbf{a}' \mathbf{R_K} \tag{11}$$

---

[4]It is also possible to impose many conditions simultaneously. For example, Cochrane and Saá-Requejo (2000) consider the no good-deal condition and the no arbitrage condition jointly. In section 2.5.1, we examine empirically if the no arbitrage condition represents a binding constraint for the SDFs considered in this paper. We find the SDFs are almost always positive. Hence, we conclude that the no arbitrage condition does not provide a meaningful empirical restriction in our sample and do not consider it further.

(12) $$w = R_{MF} - \mathbf{c}'\mathbf{R_K},$$

where:

(13) $$\mathbf{a}' = \mathbf{1}' E[\mathbf{R_K}\,\mathbf{R_K'}]^{-1}$$

(14) $$\mathbf{c}' = E[R_{MF}\,\mathbf{R_K'}]\,E[\mathbf{R_K}\,\mathbf{R_K'}]^{-1}$$

(15) $$v = \sqrt{\frac{\left(\dfrac{(1+\bar{h}^2)}{R_F{}^2} - E[m^{*2}]\right)}{E[w^2]}}.$$

We call the solution $\bar{\alpha}_{MF}$ the "best clientele alpha" to refer to our earlier discussion; it indicates whether mutual funds add value from the perspective of their best potential clienteles, the class of investors most favorable to the mutual fund. Similarly, $\bar{m}$ represents the "best clientele SDF". In this solution, m^* is the SDF identified by Hansen and Jagannathan (1991) as having minimum volatility under the law-of-one-price condition. It is a linear function of the passive portfolio returns $\mathbf{R_K}$. The error term w represents the difference between the mutual fund return R_{MF} and the best "hedging" or "replicating" payoff $\mathbf{c}'\mathbf{R_K}$ that can be obtained from passive portfolio returns. Hence, w is the part of the mutual fund return that is not spanned by passive portfolio returns. Finally, v is the parameter that accounts for the no-good-deal restriction and is a function of the maximum Sharpe ratio $\bar{h}$.

We can further economically understand the solution by rewriting it as follows:

(16) $$\bar{\alpha}_{MF} = E[m^*R_{MF}] - 1 + v\,E[w^2].$$

34

This equation shows that the best clientele alpha can be decomposed into two parts. The first part, $E[m^*R_{MF}] - 1$, is the law-of-one-price (LOP) alpha developed by Chen and Knez (1996), based on the minimum volatility SDF. Similar to the best clientele performance measure, the LOP measure gives zero performance to passive portfolios by construction and thus does not suffer from the benchmark choice problem. It has also been used by Dahlquist and Söderlind (1999), Farnsworth, Ferson, Jackson and Todd (2002), and Ahn, Cao and Chrétien (2009), among others.

The second part, $v E[w^2]$, can be viewed as the investor disagreement between the best clientele alpha and the LOP alpha. This disagreement can come from two distinct sources: the replication error w and the maximum Sharpe ratio restriction $\bar{h}$. With regard to the first source, if the passive portfolios perfectly span the mutual fund return, so that $w = 0$, then there can be no disagreement in evaluation. Otherwise, as the expected squared replication error for a fund becomes larger, such that it becomes tougher for investors to get the « same type » of opportunities from passive portfolios, then the potential disagreement among investors becomes greater. With regard to the second source, if the maximum Sharpe ratio allowed corresponds to the highest Sharpe ratio attainable in the passive portfolios, such that $\bar{h} = h^*$ and $E[\bar{m}^2] = E[m^{*2}]$, then there is no valuation disagreement because $v = 0$. In this case, no opportunity better than the ones from passive portfolios is deemed reasonable by investors. Otherwise, as the additional opportunities that investors would find admissible (*i.e.,* not consider good deals) increase, as specified by a larger difference $\bar{h} - h^*$, the potential disagreement among investors increases.

Finally, it is also possible to develop a conditional version of the best clientele performance evaluation by following the scaled payoffs strategy of Chen and Knez (1996) and Ferson, Henry and Kisgen (2006), among others. Specifically, we form public information-managed payoffs, denoted by $\mathbf{R_Z}$, by multiplying passive returns with lagged publicly available information variables, denoted by $\mathbf{Z}$. Let $\mathbf{1_Z}$ be the corresponding prices of these payoffs, obtained by multiplying the unit vector by the lagged publicly available information variables. Then, a conditional estimation of the best clientele alpha is obtained by replacing $\mathbf{R_K}$ in the previous solution with $\mathbf{R_K^A}$, an augmented set of assets that includes both $\mathbf{R_K}$ and $\mathbf{R_Z}$, and by replacing the unit vector $\mathbf{1}$ with $\mathbf{1^A}$, which contains both $\mathbf{1}$ and $\mathbf{1_Z}$.

2.3 Methodology

2.3.1 Estimation

The solution for the best clientele performance evaluation measure necessitates the estimation of $2K + 1$ parameters for $\bar{m}$ (**a**, **c**, v), along with the alpha ($\bar{\alpha}_{MF}$). These parameters can be estimated and tested for significance using Hansen's generalized method of moments (GMM; 1982). For a sample of T observations, we rely on the following $2K + 2$ moments:

$$(17) \qquad \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}})\mathbf{R_{Kt}}] - \mathbf{1} = 0,$$

$$(18) \qquad \frac{1}{T}\sum_{t=1}^{T}[(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}})\mathbf{R_{Kt}}] = 0,$$

$$(19) \qquad \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}}) + v(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}})]^2 - \frac{\left(1 + \bar{h}^2\right)}{R_F{}^2} = 0,$$

$$(20) \qquad \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}} + v(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}}))R_{MFt}] - 1 - \bar{\alpha}_{MF} = 0.$$

The K moments in equation (17) allow for the estimation of the LOP SDF, $m_t^* = \mathbf{a}'\mathbf{R_{Kt}}$, by ensuring that it correctly prices K passive portfolio returns. The K moments in equation (18) represent orthogonality conditions between the replication error term, $w_t = R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}}$, and passive portfolio returns, which are needed to estimate the coefficients **c** in the best replicating payoff $\mathbf{c}'\mathbf{R_{Kt}}$. The moment in equation (19) imposes the no-good-deal condition to estimate the parameter v, which is restricted to be positive to obtain an upper bound. In this moment, R_F represents a risk-free rate equivalent and is simply set to one plus the average one-month Treasury bill return in our sample, which is 0.3393%. For

36

consistency, we also include this one-month Treasury bill return as one of the passive portfolio returns so that the estimated mean SDF is similar to $1/R_F$. Finally, using the estimated best clientele SDF, $\overline{m}_t = m_t{}^* + vw_t$, we obtain the upper performance bound for a mutual fund using the moment specified by equation (20).

For comparison with the best clientele alpha, we also examine the LOP performance measure of Chen and Knez (1996), which is based on the SDF with the lowest volatility. Specifically, the LOP alpha can be estimated with the following additional moment:

$$(21) \qquad \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}})R_{MFt}] - 1 - \alpha_{LOP} = 0.$$

To estimate a conditional best clientele alpha that varies according to publicly available information, we can use the augmented sets of payoffs and prices defined in section 2.2.3 to replace $\mathbf{R_{Kt}}$ with $\mathbf{R_{Kt}^A}$ and $\mathbf{1}$ with $\mathbf{1_{t-1}^A}$ in equations (17), (18) and (19) and substitute the moment in equation (20) with the following moments:

$$(22) \qquad \frac{1}{T}\sum_{t=1}^{T}\left[\left(\mathbf{a}'\mathbf{R_{Kt}^A} + v\left(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}^A}\right)\right)R_{MFt}\right] - 1 - (\bar{\alpha}_{MF0} + \bar{\boldsymbol{\alpha}}_{\mathbf{MF1}}'\mathbf{Z_{t-1}}) = 0.$$

$$(23) \qquad \frac{1}{T}\sum_{t=1}^{T}\left[\left(\mathbf{a}'\mathbf{R_{Kt}^A} + v\left(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}^A}\right)\right)\mathbf{R_{MFZt}}\right] - \mathbf{1_{Zt-1}} - (\bar{\alpha}_{MF0} + \bar{\boldsymbol{\alpha}}_{\mathbf{MF1}}'\mathbf{Z_{t-1}}) = 0.$$

These moments use a scaled version of the mutual fund return, $\mathbf{R_{MFZt}} = R_{MFt} \times \mathbf{Z_{t-1}}$, with its associated price $\mathbf{1_{Zt-1}} = 1 \times \mathbf{Z_{t-1}}$, to estimate a conditional best clientele alpha that is linear in the information variables and given by $\bar{\alpha}_{MF0} + \bar{\boldsymbol{\alpha}}_{\mathbf{MF1}}'\mathbf{Z_{t-1}}$.

In all cases, the estimation system is just identified because the number of parameters is equal to the number of moments. Hence, the parameter estimates are not influenced by the choice of weighting matrix in GMM. Furthermore, although the system is estimated for one

fund at a time, Farsnworth, Ferson, Jackson and Todd (2002) show that this strategy produces the same point estimates and standard errors for alpha as estimating a system that includes an arbitrary number of funds.[5] Finally, the statistical significance of the parameters is assessed with standard errors adjusted for conditional heteroskedasticity and serial correlation using the method of Newey and West (1987) with two lags.[6]

2.3.2　Maximum Sharpe Ratio Choice

To implement the best clientele performance measure, two choices are particularly important: passive portfolios and the maximum Sharpe ratio. In the data section, we introduce three different sets of passive portfolios to assess the sensitivity of the results to this choice. This section discusses the maximum Sharpe ratio choice, which we base on the existing literature.

In general, the literature shows that researchers typically impose a subjective constraint on the maximum Sharpe ratio. One early contribution is Ross (1976). To study deviations from the arbitrage pricing theory, he imposes a maximum Sharpe ratio of twice the market Sharpe ratio, leading to a value of 0.25. Considering that the market portfolio should according to the CAPM have the highest Sharpe ratio, he argues that this restriction should reasonably account for all attainable Sharpe ratios. With a related argument that high Sharpe ratios are unlikely from the perspective of risk-based models, MacKinlay (1995) considers that a squared annual Sharpe ratio higher than approximately 0.6 is implausibly high.

In applying their no-good-deal bounds to S&P 500 option pricing, Cochrane and Saá-Requejo (2000) select the maximum Sharpe ratio by ruling out opportunities with a Sharpe ratio greater than twice that of the S&P 500 (or, equivalently, twice the Sharpe ratio of their basis asset). They explain that this choice is not definitive and can be changed. Pyo

---

[5] Ferson, Kisgen and Henry (2007) demonstrate that this result also holds in the case of a time-varying alpha.
[6] We choose two lags to control for the non-zero serial correlation in monthly returns of a significant fraction of funds, which might invest in thinly traded stocks. For example, Ferson and Chen (2015, Table 1) report a median serial correlation of 0.12, with 25% of funds having a serial correlation greater than 0.19, potentially due to microstructure effects like non synchronous trading (Lo and MacKinlay (1990)). As a robustness check, we also reproduce the results of Table 2.3 by using no lag or four lags and find similar results.

(2011) uses the same assumption in his empirical study. Huang (2013) develops an upper bound on the R-squared of predictive regressions using the no-good-deal approach and also picks twice the market Sharpe ratio as maximum ratio. Floroiu and Pelsser (2013) price real options using no-good-deal bounds and similarly calibrate their bounds using twice the Sharpe ratio of the S&P 500. A few papers consider different values of the maximum Sharpe ratio. Kanamura and Ohashi (2009) use values ranging from two to three times the value of the market Sharpe ratio to find upper and lower bounds for summer day options. Martin (2013) provides upper bounds on risk aversion using the no-good-deal approach and calibrates his bounds with three different values for the maximum annual Sharpe ratio, i.e., 0.75, 1 and 1.25.

Overall, although the maximum Sharpe ratio is somewhat subjectively specified, the literature offers some guidance on its selection. The most common choice is a maximum Sharpe ratio of twice the one of the underlying basis assets (which oftentimes include only an equity index). This choice can also be seen as adding the market Sharpe ratio to the maximum Sharpe ratio of the basis assets. In this paper, we follow this guidance by adding to the attainable Sharpe ratio of our passive portfolios a value of 0.1262, corresponding to the monthly Sharpe ratio of the market index (the CRSP value-weighted index) in our sample. We denote this choice by $\bar{h} = h^* + hMKT$. More conservatively, we also consider adding half of this value as additional allowable opportunities, so that $\bar{h} = h^* + 0.5hMKT$. Although these are our two basic choices, we also examine in section 2.2.5 the effects on the results of other sensible maximum Sharpe ratio choices, such as directly doubling the attainable Sharpe ratio of the passive portfolios.

2.3.3   Cross-Sectional Performance Statistics

To summarize the alpha estimates for mutual funds in our sample, we provide numerous cross-sectional statistics. First, we provide the mean, standard deviation and selected percentiles of the distributions of estimated alphas and their corresponding *t*-statistics, computed as $t = \hat{\alpha}/\hat{\sigma}_{\hat{\alpha}}$, where $\hat{\alpha}$ is the estimated alpha, and $\hat{\sigma}_{\hat{\alpha}}$ is its Newey-West standard error. Second, we present *t*-statistics to test the hypothesis that the cross-sectional mean of estimated alphas is equal to zero. To perform this test, we assume that the cross-sectional

distribution of alphas is multivariate normal with a mean of zero, a standard deviation equal to the observed cross-sectional standard deviation and a correlation between any two alphas of 0.044. This last value matches the average correlation between fund residuals, adjusted for data overlap, reported by Barras, Scaillet and Wermers (2010, p. 193) and by Ferson and Chen (2015, Appendix, p. 62) when discussing the cross-sectional dependence in alpha among funds in their samples (which are similar to ours).

Third, we compute the proportions of estimated alphas that are positive, negative, significantly positive at the 2.5% level and significantly negative at the 2.5% level, and we report the p-values on the significance of these proportions using the following likelihood ratio test proposed by Christoffersen (1998) based on a binomial distribution:[7]

$$(24) \qquad LR = 2Log\left[\frac{\left(1 - \frac{n}{N}\right)^{N-n}\left(\frac{n}{N}\right)^n}{(1 - pr)^{N-n}(pr)^n}\right] \sim \chi^2(1),$$

where n is the number of funds that respects a given criterion (i.e., being positive, negative, significantly positive or significantly negative), N is the total number of funds, $\frac{n}{N}$ is the empirical proportion tested, and pr is the expected probability under the null.

Fourth, to control for mutual funds that exhibit significant alphas by luck or "false discoveries", we apply the technique of Barras, Scaillet and Wermers (2010, hereafter BSW) and its extension by Ferson and Chen (2015, hereafter FC). Their ideas consist of computing the proportion of funds with t-statistics outside the thresholds implied by a significance level (γ), denoted by t^- and t^+, then removing from it the fraction of funds that exhibit large estimated alphas by luck, and finally adjusting the result to account for the power of the tests for detecting skilled or unskilled funds. The approach thus provides proportions adjusted for false discoveries.

---

[7]A first test examines whether the proportions of positive or negative alphas are equal to 50%. A second test examines whether the proportions of significantly positive alphas or significantly negative alphas are equal to 2.5%.

Using the generalization proposed by FC, the proportions of unskilled funds ($\hat{\pi}^-$), skilled funds ($\hat{\pi}^+$) and zero performance funds ($\hat{\pi}^0$) can be found by solving the following system of equations:

(25)
$$\widehat{prob}(t > t^+) = \frac{\gamma}{2} \times \hat{\pi}^0 + \delta^+ \times \hat{\pi}^- + \beta^+ \times \hat{\pi}^+$$

(26)
$$\widehat{prob}(t < t^-) = \frac{\gamma}{2} \times \hat{\pi}^0 + \beta^- \times \hat{\pi}^- + \delta^- \times \hat{\pi}^+$$

(27)
$$\hat{\pi}^0 = 1 - \hat{\pi}^+ - \hat{\pi}^-.$$

In these equations, the probability $\widehat{prob}(t > t^+)$ ($\widehat{prob}(t < t^-)$) is the proportion of funds with t-statistics greater than t^+ (less than t^-), $\frac{\gamma}{2}$ is the size of the test, δ^+ (δ^-) is the probability that an unskilled (a skilled) fund is erroneously classified as a skilled (an unskilled) fund, and β^+ (β^-) is the probability that a skilled (an unskilled) fund is correctly classified. Hence, the proportion of skilled funds $\hat{\pi}^+$ is equal to the in-sample proportion of significantly positive funds $\widehat{prob}(t > t^+)$ less the proportions of funds with significantly positive performance by luck ($\frac{\gamma}{2} \times \hat{\pi}^0$ for lucky zero performance funds and $\delta^+ \times \hat{\pi}^-$ for very lucky unskilled funds), adjusted for the test power for skilled funds β^+. The proportion of unskilled funds $\hat{\pi}^-$ has a similar interpretation with respect to significantly negative funds, and the proportion of zero performance funds $\hat{\pi}^0$ is the proportion of funds that are not classified as skilled or unskilled funds.

We follow BSW and FC in implementing the false discovery approach. As demonstrated by FC, BSW use a special case where $\delta^+ = \delta^- = 0$ and $\beta^+ = \beta^- = 1$, which can be solved as $\hat{\pi}^0 = \frac{\widehat{prob}(t^- < t < t^+)}{1 - \gamma}$, $\hat{\pi}^+ = \widehat{prob}(t > t^+) - \frac{\gamma}{2} \times \hat{\pi}^0$ and $\hat{\pi}^- = \widehat{prob}(t < t^-) - \frac{\gamma}{2} \times \hat{\pi}^0$. They advocate values of $t^- = -0.5$ and $t^+ = 0.5$ as efficient thresholds to classify

adequately funds, which corresponds to a size of 0.3085. These choices result in our BSW classification.[8]

FC argue that the special case proposed by BSW leads to biases because their assumed values for δ^+, δ^-, β^+ and β^- are unrealistic; instead, they use simulations to calibrate these parameters. Advocating a size of 0.05, which leads to asymptotic values of $t^- = -1.645$ and $t^+ = 1.645$, they obtain the following parameters (see their Table 2.2, panel A): $\delta^+ = 0.05$, $\delta^- = 0.04$, $\beta^+ = 0.604$ and $\beta^- = 0.512$. Finally, they solve the system of equations numerically by minimizing the sum of squared errors of equations (25) and (26) subject to the Kuhn-Tucker conditions for the constraints that $\hat{\pi}^+ \geq 0$, $\hat{\pi}^- \geq 0$ and $\hat{\pi}^+ + \hat{\pi}^- \leq 1$, which ensure the positivity of the proportions. These choices result in our FC classification.[9]

2.4   Data

2.4.1   Mutual Fund Returns

Our fund data consist of monthly returns on actively managed open-ended U.S. equity mutual funds from January 1984 to December 2012. Our data source is the *CRSP Survivor-Bias-Free US Mutual Fund Database*. Following Kacperczyk, Sialm and Zheng (2008), we exclude bond funds, balanced funds, money market funds, international funds and funds that are not strongly invested in common stocks to focus on U.S. equity funds. Specifically, U.S. equity funds are identified using the following four types of codes: policy codes, Strategic Insight objective codes, Weisenberger objective codes and Lipper objective

---

[8]The false discovery adjustment of BSW can lead to a negative proportion of unskilled or skilled funds when the unadjusted observed proportion is close to zero. In such instances, we follow BSW by setting the adjusted proportion to zero and readjusting the proportion of zero performance funds so that the proportions sum to one.

[9]We do not recalibrate the parameters δ^+, δ^-, β^+ and β^- with simulations as our fund sample is very similar to the FC sample. In section 2.5.6, we examine the robustness of our results to using simulated critical values at a size of 0.05 for t^- and t^+, instead of asymptotic values, and find that our conclusions are robust. Following BSW and the base case proposed by FC, our classifications assume that the true mean alphas are equal to 0.317% for skilled funds and -0.267% for unskilled funds. FC refine the approach further by estimating simultaneously the true mean alphas and proportions of skilled and unskilled funds, and by estimating simpler models with only two alpha groups. We do not pursue these refinements.

codes.[10] The four types of codes are useful because each is available for only a part of our sample period. For example, the Lipper objective codes data start from December 1999. To focus on actively managed funds, we exclude index funds identified by the Lipper objective codes SP and SPSP and funds with a name that includes the word "index". We also exclude mutual funds that are not open-ended by consulting the variable "open to investors" in the database. Finally, we keep the funds only if they hold, on average, between 80% and 105% in common stocks.

From this initial sample of funds, we make further sampling decisions to alleviate known biases in the CRSP mutual fund database. Survivorship bias is one of the most well-documented problems in mutual fund data. It occurs when only surviving funds are sampled out of a population in which some funds enter and leave. Following Fama and French (2010), we select 1984 as our starting year because the CRSP mutual fund database is free from this bias from then on. This starting year also eliminates a related selection bias in the early years of the database, as discussed by Elton, Gruber and Blake (2001) and Fama and French (2010). Back-fill and incubation biases are studied by Evans (2010). Back-fill bias arises because the database includes fund returns that are realized prior to the fund database entry. Incubation bias refers to a situation where only funds that perform well in an incubation period are eventually open to the public and included in the database. To address these biases, we follow Elton, Gruber and Blake (2001) and Kacperczyk, Sialm and Zheng (2008). We eliminate observations before the organization date of the funds, funds that do not report their organization date, and funds without a name because they tend to correspond to incubated funds. We also exclude funds that have total net assets (TNA) inferior to $15 million in the first year of entering the database.

As a last sampling choice, following Barras, Scaillet and Wermers (2010) and others, we consider a minimum fund return requirement of 60 months, given that our GMM estimation system for SDF alpha calls for enough observations to obtain reliable statistical estimates. Fama and French (2010) argue that this requirement introduces a survivorship

---

[10]As in Kacperczyk, Sialm and Zheng (2008), we identify U.S. equity funds by policy codes: CS; Strategic Insight objective codes: AGC, GMC, GRI, GRO, ING or SCG; Weisenberger objective codes: G, G-I, AGG, GCI, GRO, LTG, MCG or SCG and Lipper objective codes: EIEI, EMN, LCCE, LCGE, LCVE, MATC, MATD, MATH, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE or SCVE.

bias in the results and use instead an eight-month survival screen to estimate their alphas based on regressions. Estimating SDF alphas in our GMM setup is obviously not feasible with only eight observations. Furthermore, an important part of our analysis consists of comparing best clientele alphas with LOP alphas, and survivorship bias should affect them similarly. Nevertheless, as a robustness check, we follow Barras, Scaillet and Wermers (2010) by estimating our main performance measures for funds with at least 36 months of returns. As they do, we find that our results are similar, and our conclusions are not altered by this choice.[11]

Considering all previous steps, we have a final sample of 2786 actively managed open-ended U.S. equity mutual funds with returns for at least 60 months between 1984 and 2012.

2.4.2 Passive Portfolio Returns

The choice of basis assets imposes a trade-off between economic power (i.e., in theory, all assets available to mutual fund investors should be included) and statistical power (i.e., an econometric estimation imposes limitations on the number of assets). We select three different sets of basis assets to represent passive opportunities available to investors. Our basis assets always include the risk-free rate plus one of the three following sets: (1) ten industry portfolios, (2) six style portfolios and (3) the market portfolio. These assets have been widely used in the empirical asset pricing literature and the mutual fund performance evaluation literature to capture the cross-section of stock returns. Classifications based on industry, style or market sensitivities are also common in practice to categorize equity investments for investors. The inclusion of the risk-free rate accounts for cash positions in equity mutual funds and fixes the mean of the SDF to a relevant value (Dahlquist and Söderlind (1999)). By varying the number and type of assets included, we aim to examine the sensitivity of our results to these choices in light of the aforementioned trade-off.

---

[11] Specifically, we reproduce the results of Table 2.3 by using a 36-month screen rather than a 60-month screen. We find that, when using a 36-month screen, the mean LOP alpha is reduced by 0.0056% and the mean best clientele alphas are reduced by 0.0037% for a maximum Sharpe ratio of $h^* + 0.5hMKT$ and by 0.0031% for a maximum Sharpe ratio of $h^* + hMKT$.

Ten industry portfolios are used for our main results, and the data are from Kenneth R. French's website. The portfolios consist of consumer nondurables, consumer durables, manufacturing, energy, high technology, telecommunication, shops, healthcare, utilities, and other industries. Six style portfolios are obtained also from Kenneth R. French's website. The portfolios are constructed from two market equity capitalization (size) sorts (large or small) and three book-to-market (value) sorts (low, medium or high). The market portfolio is the CRSP value-weighted index. Its returns and the risk-free rate returns are taken from the CRSP database.

2.4.3   Information Variables

For conditional performance evaluation, we consider the lagged values of four public information variables that are commonly used in the literature and were first introduced by Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988) and Fama and French (1989). We use the dividend yield of the S&P 500 Index (DIV) from the Datastream database, which is computed as the difference between the log of the twelve-month moving sum of dividends paid on the S&P 500 and the log of its lagged value; the yield on three-month U.S. Treasury bills (YLD) from the FRED database at the Federal Reserve Bank at St. Louis; the term spread (TERM), which is the difference between the long-term yield on government bonds (from Datastream) and the yield on the three-month Treasury bills; and the default spread (DEF), which is the difference between BAA- and AAA-rated corporate bond yields from the FRED database.

With these lagged information variables, we construct four public information-managed payoffs by combining them with the market portfolio returns. We then add these four managed payoffs to each set of basis assets described previously to obtain the augmented sets $\mathbf{R_K^A}$ used for conditional performance evaluation.

2.4.4   Summary Statistics

Table 2.1 presents the summary statistics for the monthly returns of our sample of actively managed open-ended U.S. equity mutual funds (panel A) and for the monthly returns of the basis assets and the values of the information variables (panel B). Panel A also includes

summary statistics for the SDF alphas from Carhart's (1997) model and their corresponding *t*-statistics, which will serve as a basis for comparison for the LOP and best clientele alphas.[12]

In panel A, monthly equity fund average returns (net of fees) have a mean of 0.73% and a standard deviation of 0.3% across funds. Average returns range from -4.83% to 2.09%, and standard deviations range from 0.92% to 16.92%. Monthly Sharpe ratios vary from -0.464 to 0.379, with a mean of 0.086 and a standard deviation of 0.053. The cross-sectional distribution of Carhart SDF alphas has a mean of -0.12% (*t*-stat. = -2.32) and a standard deviation of 0.246%, and 73% of funds have a negative Carhart alpha. At the 5% significance level, approximately 15% (2%) of funds have significantly negative (positive) Carhart alphas. These results are typical compared with the findings in the mutual fund literature.

In panel B, industry portfolios and style portfolios have mean monthly returns of approximately 1%. Industry portfolios have mean returns between 0.83% and 1.17% and standard deviations between 3.99% and 7.22%. Style portfolios have mean returns between 0.80% and 1.22% and standard deviations between 4.58% and 6.76%. Sharpe ratios vary from 0.070 to 0.196 for industry portfolios and from 0.069 to 0.149 for style portfolios. Statistics for the market portfolio returns, the risk-free returns and the information variables are as expected. To illustrate the investment opportunities captured by the basis assets, Figure 2.1 shows the efficient frontiers of returns from the sets based on industry portfolios (RF + 10I), based on style portfolios (RF + 6S), and based on the market portfolio (RF + MKT). As expected, the market portfolio set provides fewer investment opportunities than the other sets.

---

[12] The Carhart SDF is a linear function of the market factor, size (SMB) factor, value (HML) factor and momentum factor available on Kenneth R. French's website. For each mutual fund, we estimate jointly the parameters of the Carhart SDF and the corresponding alpha by using GMM with a just identified system. Specifically, we estimate the five parameters of the Carhart SDF by requiring the SDF to price correctly the one-month Treasury bill return and the four Carhart factors, and the alpha by using a moment similar to equation (21), but with the Carhart SDF replacing the LOP SDF.

2.5 Empirical Results

2.5.1 Best Clientele Stochastic Discount Factors

We begin our empirical analysis by examining the estimated best clientele SDFs and the sources of disagreement. As discussed in section 2.2.3, the best clientele SDF $\bar{m}$ depends on the error term w, related to the ability of passive portfolios to span mutual fund returns, and the disagreement parameter v, which accounts for the no-good-deal restriction and is a function of the maximum Sharpe ratio. Table 2.2 reports cross-sectional statistics on R-squared from regressions of fund returns on different sets of passive portfolios and, for comparison, on Carhart's (1997) factors (panel A), on estimates of the disagreement parameter v (panel B) and on SDFs used for performance evaluation (panel C). Panels B and C show the results using the risk-free rate and ten industry portfolios as passive portfolios.

Panel A is informative on the ability of passive portfolios to replicate mutual fund payoffs. A higher R-squared should tighten the bound associated with the best clientele alpha. The panel shows that the six style portfolios provide the highest mean R-squared at 85.3%, followed by the Carhart factors at 83.5% and the industry portfolios at 81.6%. The market portfolio provides the lowest mean R-squared at 74.8%. Although the three sets are similar in their 99th percentile R-squared, industry portfolios are better at spanning the fund returns that are most difficult to replicate, as indicated by the R-squared at the first percentile. Panel B shows that best clientele SDFs include a significant disagreement parameter v in most cases. When the maximum Sharpe ratio $\bar{h} = h^* + 0.5hMKT$, the estimates of v have a mean of 11.0, with t-statistics averaging 1.8. They are statistically significant at the 10% level for 67.4% of the funds. When $\bar{h} = h^* + hMKT$, the estimates have a mean of 13.8, with statistically significant values for all funds. Panel C looks at the empirical SDFs used for performance evaluation in two best clientele measures and in the LOP measure (denoted by h^*). As expected, given that the risk-free rate is part of the basis assets, the three SDFs have the same mean. Consistent with the increased estimates of v, the average across funds of the SDF standard deviations increases with the maximum Sharpe ratio. However, the higher volatility does not result in SDFs having undesirable

economic properties. In particular, the proportions of negative SDFs are well below 1% in all cases, suggesting that imposing the no-arbitrage condition to ensure the positivity of SDFs would not materially affect our results.

2.5.2   Best Clientele Performance Results

Table 2.3 presents the main empirical results. Using the risk-free rate and ten industry portfolios as basis assets, it shows statistics on the cross-sectional distribution of SDF alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $\bar{h} = h^* + 0.5hMKT$ and $\bar{h} = h^* + hMKT$. The results for the LOP measure of Chen and Knez (1996) (denoted by h^*) are also reported for comparison. Figure 2.2 illustrates these results by presenting histograms on the distributions of LOP alphas and either best clientele alphas for $\bar{h} = h^* + 0.5hMKT$ (Figure 2.2a) or best clientele alphas for $\bar{h} = h^* + hMKT$ (Figure 2.2b).

In panel A of Table 2.3, we provide the mean, standard deviation and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). We also report the t-statistics on the significance of the cross-sectional mean of the estimated alphas (see t-stat). As discussed in section 2.3.3, the test accounts for the cross-sectional dependence in performance among funds by assuming a correlation between any two alphas of 0.044.

The distribution of SDF alphas from the best clientele measure with $\bar{h} = h^* + hMKT$ has a mean of 0.444% and a standard deviation of 0.418%. When $\bar{h} = h^* + 0.5hMKT$, the mean and standard deviation decrease to 0.236% and 0.334%, respectively. Both means are significantly different from zero, with respective t-statistics of 3.35 and 5.04. For comparison, the average alpha from the LOP measure, which does not attempt to capture the evaluation for best clienteles by ruling out investor disagreement (as $v = 0$), is -0.179% (t-stat. = -3.14). This negative performance is similar to the Carhart results presented in Table 2.1, which also do not consider investor disagreement by focusing on a unique linear factor SDF for evaluation. As in Ferson and Lin (2014), these results support an economically important divergence in performance evaluation between clienteles. For

example, the magnitude of average disagreement between the LOP alpha and the best clientele alpha with $\bar{h} = h^* + 0.5hMKT$ is 0.415%. This value is comparable to the magnitude of investor disagreement documented by Ferson and Lin (2014, Table III), who obtain bounds between 0.21% and 0.38% for the expected disagreement with traditional regression alphas for various benchmark returns. The divergence in alphas is well illustrated by the alpha distributions in Figure 2.2.

The distributions of the t-statistics confirm that the increased performance associated with the best clientele measures result in more significantly positive alphas and fewer significantly negative alphas. Panel B studies this issue further. It gives proportions of estimated alphas that are positive ($\%\bar{\alpha}_{MF} > 0$), negative ($\%\bar{\alpha}_{MF} < 0$), significantly positive ($\%\bar{\alpha}_{MF}signif > 0$), and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$). It also provides proportions adjusted for false discoveries with the techniques of Barras, Scaillet and Wermers (2010; denoted by the BSW classification) and of Ferson and Chen (2015; denoted by the FC classification), i.e., proportions of unskilled funds ($\hat{\pi}^-$), skilled funds ($\hat{\pi}^+$) and zero performance funds ($\hat{\pi}^0$). It finally presents the p-values for the likelihood ratio tests (described in section 2.3.3), which show that the proportions of positive alphas are equal to 50% and that the proportions of significantly positive and significantly negative alphas are equal to 2.5%.

The results in panel B show that the proportions of positive alphas and significantly positive alphas increase when considering best clienteles. They go from 20.32% to 91.49% for positive alphas and from 1.04% to 47.52% for significantly positive alphas. Accordingly, the proportions of negative and significantly negative alphas decrease, from 79.68% to 8.51% for negative alphas and from 29.54% to 0.47% for significantly negative alphas. The p-values confirm the significance of these results. Furthermore, the proportions of skilled funds increase considerably for best clienteles (going from 0.00% to 65.34% for the FC classification and from 0.00% to 77.94% for the BSW classification). Inversely, the proportions of unskilled funds for best clienteles disappear.[13] Again, accounting for

---

[13] For the FC classification, results for the LOP measure, which does not consider investor disagreement, are similar to results of Ferson and Chen (2015) based on the Fama-French three-factor model. For the BSW classification, they show fewer zero alpha funds and more unskilled funds compared to results of Barras,

investor disagreement and focusing on the best potential clienteles are keys to understanding the difference between our results and the existing literature on the value added by fund managers. For example, the findings from the LOP measure are comparable to the results from the Carhart SDF and typical of the literature, with less than 30% (more than 70%) of funds having positive (negative) values. Notably, increasing allowable opportunities by half the market Sharpe ratio is sufficient to obtain approximately the opposite result. This is consistent with Ahn, Cao and Chrétien (2009), who argue that more than 80% of mutual funds could be given a positive performance value by some investors.

To gain more insight into the best clientele performance evaluation, Figure 2.3 presents the best clientele and LOP alphas for decile portfolios of the 2786 mutual funds sorted in increasing order of their average return (Figure 2.3a), their standard deviation of returns (Figure 2.3b), and their Sharpe ratio (Figure 2.3c). Figures 2.3a and 2.3c show that the alpha increases with the average return and the Sharpe ratio. Thus, not surprisingly, funds with higher average returns or Sharpe ratios are generally given higher best clientele or LOP alphas. Figure 2.3b reveals that best clientele alpha also increases with the standard deviation of returns, particularly for the five decile portfolios with the highest volatility, a relation not observed for LOP alpha. Hence, the difference between best clientele and LOP alphas (a disagreement measure in the spirit of the expected disagreement with a traditional alpha proposed by Ferson and Lin (2014)) is relatively stable across portfolios formed on the average return and the Sharpe ratio, but it increases across portfolios formed on the standard deviation of returns. Intuitively, highly volatile mutual funds represent somewhat "unique" opportunities for investors because they cannot be easily replicated by passive portfolio returns. This "uniqueness" allows for a greater valuation disagreement between investors.

Overall, we find that an increase in admissible investment opportunities equivalent to half the Sharpe ratio of the market index leads to a generally positive performance for best clienteles. As stipulated by Chen and Knez (1996) and Ferson and Lin (2014), if the

---

Scaillet and Wermers (2010) based on the Carhart model. This finding is consistent with the distribution of t-statistics being more to the left for LOP alphas than for Carhart alphas, combined with the analysis of Ferson and Chen (2015) that finds that the BSW classification is sensitive to the choice of test size.

SDF alpha is positive, then there are some investors who would want to buy the fund. Our best clientele alpha results suggest that there are clienteles who would want to buy a majority of mutual funds, consistent with the real-life continued investments in these vehicles.

2.5.3 Conditional Best Clientele Performance Results

A large body of literature, starting with Ferson and Schadt (1996) and Christopherson, Ferson and Glassman (1998), argues that accounting for public information results in improved performance measures and that alpha varies across the business cycle. In particular, Glode (2011) shows that mutual funds could be valuable to their clienteles by providing positive alphas during recessions, when their marginal utility (or SDF) is high. Moskowitz (2000), Kosowski (2011) and Kacperczyk, Van Niewerburgh and Veldkamp (2014) also find some evidence of better mutual fund performance in recessions. We implement a conditional version of our performance measure that considers the best clienteles of a mutual fund in an incomplete market with investor disagreement. To do so, we use the risk-free rate, ten industry portfolios and the public information-managed payoffs described in section 2.4.3 to form an augmented set of basis assets, and we take the system of moments, including equations (22) and (23), for estimation purposes.

Table 2.4 presents the results for the conditional version of best clientele alphas. To revisit the findings of Glode (2011) and others, it gives statistics on average conditional alphas and average conditional alphas in expansions and recessions, with months classified according to the *NBER US Business Cycle Expansions and Contractions Reference Dates*. The table shows that the unconditional findings of the previous section extend to the average conditional results. By analyzing differences between Tables 3 and 4, we find that the conditional version decreases the alpha for 53% of funds using the LOP measure and for 74% and 80% of the funds using the best clientele performance measures, with $\bar{h} = h^* + 0.5hMKT$ and $\bar{h} = h^* + hMKT$, respectively. However, the performance changes are less than five basis points for more than 90% of the funds.

Notably, the results in expansions versus recessions are generally consistent with the findings of Glode (2011). Although the mean values are similar, it becomes apparent

that alphas are more positive in recessions than in expansions when comparing median values or looking at the proportions in Panel B. For example, the best clientele measure with $\bar{h} = h^* + 0.5hMKT$ provides conditional alphas with, respectively, a mean and median of 0.220% and 0.168% in expansions, versus 0.260% and 0.400% in recessions. Its proportions of significantly positive alphas are 30.3% in expansions versus 49.9% in recessions, and its proportions of skilled funds with the FC classification are 36.2% in expansions versus 72.3% in recessions. Overall, the inclusion of conditioning information does not alter our conclusion on the importance of investor disagreement and best clienteles. We find a generally positive performance for best clienteles, with evidence that it is more favorable in recessions than in expansions.

2.5.4 Sensitivity to Passive Portfolio Choice

Tables 4 and 5 allow for an examination of the result sensitivity to the choice of basis assets. They show unconditional performance results using basis assets based on six style portfolios (Table 2.5) and the market portfolio (Table 2.6). In the latter case, the LOP measure is equivalent to the CAPM measure because the SDF is linear in the market return, $m^* = a_1 R_F + a_2 R_{MKT}$.

The previous findings are confirmed when using these alternative sets of basis assets. An increase in admissible investment opportunities equivalent to half the market Sharpe ratio leads to generally positive best clientele performance values, and more skilled funds than unskilled funds, for all sets of basis assets. For example, alphas estimated from the best clientele performance measure with $\bar{h} = h^* + 0.5hMKT$ have a mean of 0.289% (t-stat. = 3.98) for the six style portfolio set and 0.270% (t-stat. = 4.23) for the market portfolio set. These values are slightly greater than the mean of 0.236% for the ten industry portfolio set. This result, along with a general comparison of the cross-sectional distributions of alphas from the three sets, suggests that the benchmarks implicit in the six style portfolios or the market portfolio appear slightly easier to "beat" on a risk-adjusted basis.

As before, the means of the SDF alpha distributions indicate an economically important divergence in performance evaluation between clienteles. For example, the

magnitudes of average disagreement between LOP alphas and best clientele alphas with $\bar{h} = h^* + 0.5hMKT$ are comparable across different basis assets (i.e., 0.415% for the ten industry portfolio set, 0.373% for the six style portfolio set and 0.339% for the market portfolio set). The higher disagreement for the ten industry portfolio set has two sources. Compared to style portfolios, industry portfolios span fund returns slightly less well, as shown in panel A of Table 2.2. Compared to the market portfolio, industry portfolios result in higher maximum Sharpe ratios, and hence higher disagreement parameters, due to their higher attainable Sharpe ratio h^*.

2.5.5 Alternative Maximum Sharpe Ratios

Table 2.7 presents empirical results for other sensible choices of maximum Sharpe ratios, using the risk-free rate and ten industry portfolios as basis assets. As discussed earlier, several papers argue that the maximum Sharpe ratio $\bar{h}$ is a subjective choice. We explore three additional approaches for setting $\bar{h}$. In the first approach, we select it as a multiple of the attainable Sharpe ratio of the passive portfolios. Specifically, we consider $\bar{h} = 1.5h^*$ and $\bar{h} = 2h^*$. This approach is in line with the previously reviewed literature that chooses twice the Sharpe ratio of the basis assets. However, the sample h^* can be near zero or unusually high, particularly for funds with a limited time series. Taking a multiple of a potentially unrealistic h^* might lead to an unrealistic $\bar{h}$. In the second approach, we thus *add* to h^* a fraction of the full-sample optimal basis asset Sharpe ratio. The maximum Sharpe ratios become $\bar{h} = h^* + 0.5hT$ and $\bar{h} = h^* + hT$, where hT represents the optimal Sharpe ratio of the basis assets in the full sample. In the third approach, because the sample hT might be biased upward, we use instead an adjusted Sharpe ratio hTa following the bias correction proposed by Ferson and Siegel (2003).[14] The maximum Sharpe ratios are then $\bar{h} = h^* + 0.5hTa$ and $\bar{h} = h^* + hTa$.

---

[14] Ferson and Siegel (2003) show that the sample optimal Sharpe ratio is biased upward when the number of basis assets (K) is large relative to number of observations (T). Their proposed correction is $hTa = \sqrt{\frac{(hT)^2 \ (T-K-2)}{T} - \frac{K}{T}}$.

The empirical results in Table 2.7 show that SDF alphas estimated from the best clientele performance measure have means varying from 0.297% (t-stat. = 3.95) for $\bar{h} = h^* + 0.5hTa$ to 0.797% (t-stat. = 6.34) for $\bar{h} = 2h^*$. All maximum Sharpe ratios investigated lead to best clientele performance values that are generally positive and increasing with the importance of additional opportunities allowed by the choice of $\bar{h}$. Average investor disagreement, computed as the difference between the mean alpha in Table 2.7 and the mean LOP alpha in Table 2.3 (under h^*), continues to be economically important. For example, when $\bar{h} = h^* + 0.5hTa$, we obtain an average disagreement of 0.476%. Appendix 2.A documents similar findings using basis assets based on six style portfolios or the market portfolio. Overall, these results show that the maximum Sharpe ratio of $\bar{h} = h^* + 0.5hMKT$, investigated in previous sections, is a relatively conservative choice. It adds fewer investment opportunities than the other sensible maximum Sharpe ratios that can be justified from the literature.

2.5.6 Finite Sample Properties of Best Clientele Alphas

All previous results use the asymptotic GMM theory of Hansen (1982), along with Newey and West's (1987) standard errors, to make inferences on estimated alphas. However, as first documented by Ferson and Foerster (1994) in an asset-pricing context, the finite sample properties of GMM estimators can deviate from their asymptotic properties. For mutual funds, Kosowski, Timmermann, Wermers and White (2006) and Fama and French (2010) are examples of studies on the finite sample properties of regression-type alpha estimates. This section provides finite sample evidence on our SDF alpha estimates by conducting bootstrap simulations.

Specifically, we conduct a bootstrap experiment that imposes the null hypothesis that alpha is zero by adapting the procedure proposed in Fama and French (2010) and Ferson and Chen (2015) to the case of SDF alpha. First, we create adjusted (gross) mutual fund returns, defined as $R_{MFt}^{Adj} = R_{MFt} - \widehat{\alpha}_{MF}/\widehat{E(m)}$, where $\widehat{\alpha}_{MF}$ is the best clientele or LOP alpha estimate in the actual data, and $\widehat{E(m)}$ is the mean in the actual data of the best clientele or LOP SDF associated with $\widehat{\alpha}_{MF}$. Using these adjusted mutual fund returns in the simulations imposes the null that the "true" alphas are zero. Second, for all funds, we form

a simulated sample with a size equal to the total number of observations by drawing with replacement from their adjusted returns and the passive portfolio returns. Each draw picks the data that correspond to a randomly selected date, hence capturing the correlations across funds. This bootstrap procedure is repeated to create 1000 samples. Following Ferson and Chen (2015), we apply the 60-month survival screen only after a fund is drawn for an artificial sample.[15] Third, we obtain the empirical distributions of the SDF alpha t-statistics by computing the alpha estimates and their t-statistics for each of the 1000 samples, following the estimation strategy of section 2.3.1.

Table 2.8 presents, in the case of the passive portfolios based on ten industry portfolios, the results of using bootstrap empirical distributions of the SDF alpha t-statistics for inference purposes. Panel A gives statistics on the distributions of bootstrap p-value statistics for the LOP measure and two best clientele performance measures (with a maximum Sharpe ratio of either $\bar{h} = h^* + 0.5hMKT$ or $\bar{h} = h^* + hMKT$). Panel B presents proportions of alphas that are significantly positive ($\%\bar{\alpha}_{MF}signif > 0$) and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$) using the bootstrap p-values, and proportions adjusted for false discoveries based on the simulated critical values for t-statistics that correspond to the size used in the BSW and FC classifications. Overall, the findings of previous sections on the generally positive performance for best clienteles are robust to finite sample issues. For example, the proportions in panel B, computed from simulated empirical distributions of the t-statistics, are similar to the proportions in panel B of Table 2.3, where significance is assessed with the asymptotic distribution.

2.5.7   Zero-Alpha Implied Maximum Sharpe Ratios

Our analysis has thus far relied on an exogenous choice for the maximum Sharpe ratio. Previously, we showed that although this choice is somewhat subjectively specified, the literature offers some guidance, providing a justification for our selections. This section

---

[15] The simulation procedure leads to missing values being distributed randomly in the artificial sample, while they occur mainly in blocks in the original data. As discussed by Ferson and Chen (2015), it preserves the important cross-sectional dependence between funds, but not the small serial dependence in the data. Consequently, the method of Newey and West (1987) with no lag is used for standard errors in the simulations.

investigates an alternative estimation strategy that does not require the selection of a maximum Sharpe ratio and leads to an evaluation of the investor disagreement needed for mutual funds to be fairly priced by their potentially best clienteles.

To understand this strategy, notice that the selection of the maximum Sharpe ratio $\bar{h}$ allows for the estimation of the disagreement parameter v in equation (19). Then, v is needed for the estimation of $\bar{\alpha}_{MF}$ with equation (20). In this section, we proceed in reverse. Specifically, we set a value for alpha that implies that the best clientele gives zero value to a mutual fund, $\bar{\alpha}_{MF} = 0$. This choice leads to the estimation of v with equation (20), which then allows for the estimation of $\bar{h}$ with equation (19). We call the resulting estimated $\bar{h}$ the "zero-alpha implied maximum Sharpe ratio". Finally, the difference between this value for a fund and the corresponding optimal basis asset Sharpe ratio for its sample, $\bar{h} - h^*$, gives the increase in admissible investment opportunities sufficient to achieve zero alpha. As discussed in section 2.2.3, a greater difference of $\bar{h} - h^*$ leads to a wider potential valuation disagreement among investors (with no disagreement when $\bar{h} - h^* = 0$). Hence, a small value for $\bar{h} - h^*$ indicates that little investor disagreement is needed to find clienteles who give a nonnegative value to a fund.

Table 2.9 shows the distributions of the implied Sharpe ratios estimated when fixing the best clientele alpha at zero (under $\bar{\alpha}_{MF} = 0$), the attainable optimal Sharpe ratios of the passive portfolios (under Basis Assets) and the fund-by-fund differences between the Sharpe ratios. For all three sets of basis assets, the average differences are small, so only a small increase in admissible opportunities is needed to change the negative mean LOP alphas into zero mean best clientele alphas. For example, when considering ten industry portfolios as basis assets, augmenting Sharpe ratio opportunities by only 0.041 (approximately one third of the sample market Sharpe ratio) is sufficient for fund evaluation to become zero on average. Even fewer additional opportunities are needed for the passive portfolios based on six style portfolios (0.030) and the market index (0.034). The distributions show that funds require different levels of investor disagreement to be valued fairly. For example, with the basis assets based on ten industry portfolios and according to the 1st and 99th percentiles, the zero-alpha implied Sharpe ratios vary from 0.149 to 0.751, and the Sharpe ratio differences vary from 0.000 to 0.460. Nevertheless,

these findings suggest that our conclusion on the generally positive performance values for best potential clienteles would hold unless an unreasonably low value for the maximum Sharpe ratio is selected.

2.5.8 Worst Clientele Performance Results and Total Disagreement

In all previous sections, the results focus on the best clientele alpha, a useful measure for managers who hope to cater to the right clienteles and for researchers who wish to understand whether there are some investors who would want to buy a fund. In this section, we examine the "worst clientele alpha" by estimating the lower bound on the expected alpha in our setup, and we investigate the total performance disagreement found by comparing the best clientele alpha with the worst clientele alpha. The worst clientele alpha is helpful in understanding whether there are clienteles that value funds more negatively than previous evidence shows. More optimistically, a positive worst clientele alpha indicates that all investors favorably value a fund. The measure of total performance disagreement extends the results of Ferson and Lin (2014), who argue that there is an economically important divergence in performance evaluation between clienteles.

It is straightforward to obtain the worst performance alpha. Cochrane and Saá-Requejo (2000) show that the solution to the lower-bound problem can be found using the SDF given by $\underline{m} = m^* - vw$, which we call the "worst clientele SDF". Then, the worst performance alpha is given by $\underline{\alpha}_{MF} = E[\underline{m}R_{MF}] - 1$ and can be estimated empirically by GMM using the estimation strategy presented previously. Table 2.10 presents the cross-sectional performance statistics for the estimates of worst clientele alphas. The results in panel A show that the distribution of worst clientele SDF alphas with $\bar{h} = h^* + hMKT$ has a mean of -0.801% (t-stat. = -7.94) and a standard deviation of 0.480%. When $\bar{h} = h^* + 0.5hMKT$, the mean and standard deviation decrease to -0.594% (t-stat. = -7.26) and 0.388%, respectively. Both distributions clearly show more underperforming funds compared to LOP alphas or Carhart alphas (reported in Table 2.1). In fact, the proportions in panel B indicate that fewer than 1% of funds have positive worst clientele alphas, suggesting that almost all funds are negatively evaluated by some investors. From the point

of view of worst clienteles, more than 90% of mutual funds appear unskilled. Thus, the evidence from worst clienteles is more negative than that from standard models.

With estimates of best and worst clientele alphas, it is straightforward to obtain the total performance disagreement. Cochrane and Saá-Requejo (2000) analyze the difference between their upper and lower bounds. In our context, we can interpret the difference between best and worst clientele alphas, $\bar{\alpha}_{MF} - \underline{\alpha}_{MF} = 2vE(w^2)$ as a measure of total performance disagreement between mutual fund investors. This measure is different from the bound on the expected disagreement with the traditional alpha of Ferson and Lin (2014). Their bound provides an estimate of the disagreement that best or worst clienteles could have with a traditional alpha but not the larger total disagreement that they have with each other. Focusing on the best and worst clientele measures with a maximum Sharpe ratio of $\bar{h} = h^* + 0.5hMKT$, Table 2.11 shows statistics on the cross-sectional distributions of total disagreement values and their t-statistics for the three sets of basis assets. The results extend the evidence of Ferson and Lin (2014) on the economically and statistically significant performance disagreement in mutual fund evaluation. For the three sets of basis assets, total disagreement values are significantly positive for more than 99% of funds, with mean values of 0.830% (t-stat. = 8.20) for the industry portfolio set, 0.746% (t-stat. = 8.48) for the style portfolio set and 0.677% (t-stat. = 8.62) for the market portfolio set. These results are comparable with twice the values of investor disagreement documented by Ferson and Lin (2014, Table III), consistent with interpreting their bound as the expected disagreement from a traditional alpha applicable to both best and worst clienteles.[16]

2.6   Conclusion

In this paper, we apply the no-good-deal approach of Cochrane and Saá-Requejo (2000) to measure mutual fund performance from the point of view of the most favorable clienteles. This approach allows us to consider investor disagreement in mutual fund performance

---

[16] In unreported estimations, following Ferson and Lin (2014), we separate out retail funds from institutional funds on the assumption that individuals are more heterogeneous than institutional investors. We do not find any conclusive evidence of differences in best clientele alpha, worst clientele alpha or total disagreement between the two subsamples. In their analysis of fund flows and disagreement, Ferson and Lin (2014) also do not find significant differences between these classes of funds.

measurement. We use a large cross-section of actively managed open-ended U.S. equity mutual funds to provide the first comprehensive performance evaluation exercise from the point of view of the potentially most favorable clienteles of each fund.

Our empirical results suggest that the long-standing issue of actively managed mutual fund underperformance might be due to the implicit use of unique representative investors in standard performance measures. Considering investor disagreement and focusing on best clienteles cause mutual funds to perform better, with the cross-sectional average of estimated alphas increasing with additional admissible investment opportunities in an incomplete market. These results are robust to the use of different passive portfolios, different maximum Sharpe ratios (which control the investor disagreement parameter), conditioning information, and finite sample issues and adjustments for false discoveries. Overall, they support the findings of Ahn, Cao and Chrétien (2009) and Ferson and Lin (2014) on the importance of investor disagreement in mutual fund evaluation.

The best clientele performance approach can be extended in numerous ways. First, it can be interesting to identify and characterize valuable investor clienteles by studying the implied marginal preferences reflected in the best clientele SDFs. Second, it is possible to investigate how the best clientele evaluations differ from the evaluations of representative investors implicit in commonly used performance models. Such a comparison, by examining how a representative investor can misrepresent the value of funds for some of their clienteles, could form the basis of an evaluation of the appropriateness of standard measures for the purpose of performance evaluation. Finally, additional or different restrictions can be imposed on the set of investor SDFs (such as the maximum gain-loss ratio condition of Bernardo and Ledoit (2000)) to adapt the approach to measuring the performance of portfolios with nonlinear payoffs, such as hedge funds.

2.7 References

Ahn, D., Cao, H. H., Chrétien, S., (2009), "Portfolio performance measurement: a no arbitrage bounds approach"*, European Financial Management*, 15, 298-339.

Bailey, W., Kumar, A., Ng, D., (2011), "Behavioral biases and mutual fund investors", *Journal of Financial Economics*, 102, 1-27.

Barras, L., Scaillet, O., Wermers, R., (2010), "False discoveries in mutual fund performance : measuring luck in estimated alphas", *Journal of Finance*, 65, 179-216.

Bernardo, A. E., Ledoit, O., (2000), "Gain, loss, and asset pricing", *Journal of Political Economy*, 108, 144-172.

Campbell, J. Y., (1987), "Stock returns and the term structure", *Journal of Financial Economics*, 18, 373-399.

Campbell, J. Y., Shiller, R. J., (1988), "The dividend price ratio and expectations of future dividends and discount factors", *Review of Financial Studies*, 1, 195-228.

Carhart, M. M., (1997), "On persistence in mutual fund performance" *Journal of Finance*, 52, 57-82.

Chen, Z., Knez, P. J., (1996), "Portfolio performance measurement: theory and applications", *Review of Financial Studies*, 9, 511-555.

Christoffersen, P., (1998), "Evaluating interval forecasts", *International Economic Review*, 39, 841-862.

Christophersen, J. A., Ferson, W. E., Glassman, D. A., (1998), "Conditioning manager alpha on economic information: Another look at the persistence of performance" *Review of Financial Studies*, 11, 111-142.

Cochrane, J., Saá-Requejo, J., (2000), "Beyond arbitrage: good-deal asset price bounds in incomplete markets", *Journal of Political Economy,* 108, 79-119.

Cremers, M., Petajisto, A., Zitzewitz, E., (2013), "Should benchmark indices have alpha? Revisiting performance evaluation ", *Critical Finance Review,* 2, 1-48.

Dahlquist, M., Söderlind, P., (1999), "Evaluating portfolio performance with stochastic discount factors", *Journal of Business*, 72, 347-383.

Del Guercio, D., Reuter, J., (2014), "Mutual fund performance and the incentive to generate alpha", *Journal of Finance,* 69, 1673-1704.

Elton, E. J., Gruber, M. J., Blake, C. R., (2001), "A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases", *Journal of Finance*, 56, 2415–2430.

Evans, R. B., (2010), "Mutual fund incubation", *Journal of Finance*, 65, 1581-1611.

Fama, E. F., (1998), "Market efficiency, long-term returns and behavioral finance", *Journal of Financial Economics*, 49, 283-306.

Fama, E. F., French, K. R., (1989), "Business conditions and expected returns on stocks and bonds", *Journal of Financial Economics*, 25, 23-49.

Fama, E. F., French, K. R., (2010), "Luck versus skill in the cross-section of mutual fund returns", *Journal of Finance*, 65, 1915-1948.

Farnsworth, H., Ferson, W., Jackson, D., Todd, S., (2002), "Performance evaluation with stochastic discount factors", *Journal of Business*, 75, 473-503.

Ferson, W. E., (2010), "Investment performance evaluation", *Annual Review of Financial Economics*, 2, 207-234.

Ferson, W. E., Foerster, S. R., (1994), "Finite sample properties of the Generalized Method of Moments in tests of conditional asset pricing models", *Journal of Financial Economics*, 36, 29-55.

Ferson, W. E., Henry, T. R., Kisgen, D. J., (2006), "Evaluating government bond fund performance with stochastic discount factors", *Review of Financial Studies*, 19, 423-456.

Ferson, W. E., Kisgen, D. J., Henry, T. R., (2007), "Fixed income fund performance across economic states" *Research in Finance*, 23, 1-62.

Ferson, W. E., Lin, J., (2014), "Alpha and performance measurement: the effect of investor disagreement and heterogeneity", *Journal of Finance*, 69, 1565-1596.

Ferson, W. E., Siegel, A. F., (2003), "Stochastic discount factor bounds with conditioning information", *Review of Financial Studies*, 16, 567-595.

Ferson, W. E., Chen, Y., (Janury 2015), "How many good and bad fund managers are there, really?" Unpublished working paper, University of Southern California.

Floroiu, O., Pelsser, A., (2013), "Closed-form solutions for options in incomplete markets", Unpublished working paper, Maastricht University.

Glode, V., (2011), "Why mutual funds underperform." *Journal of Financial Economics*, 93, 546-559.

Glosten, L., Jagannathan, R., (1994), "A contingent claims approach to performance evaluation", *Journal of Empirical Finance,* 1, 133-66.

Gruber, M. J., (1996), "Another puzzle: The growth in actively managed mutual funds" *Journal of Finance*, 51, 783-810.

Hansen, L. P., (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029–1054.

Hansen, L. P., Jagannathan, R., (1991), "Implication of security market data for models of dynamic economies", *Journal of Political Economy*, 99, 225-262.

Huang, D., (2013), "What is the maximum return predictability permitted by asset pricing models?", Unpublished working paper, Singapore Management University.

Jensen, M. C., (1968), "The performance of mutual funds in the period 1945-1964", *Journal of Finance*, 23, 389–416.

Kacperczyk, M., Sialm, C., Zheng, L., (2008), "Unobserved actions of mutual funds", *Review of Financial Studies*, 21, 2379-2416.

Kacperczyk, M., Van Niwuwerburgh, S., Veldkamp, L., (2014), "Time-varying fund manager skill" *Journal of Finance*, 69, 1455-1484.

Kanamura, T., Ohashi, K., (2009), "Pricing summer day options by good-deal bounds", *Energy Economics*, 31, 289-297.

Keim, D. B., Stambaugh, R. F., (1986), "Predicting returns in the stock and bond markets", *Journal of Financial Economics*, 17, 357-390.

Koski, J. L., Pontiff, J., (1999), "How Are Derivatives Used? Evidence from the Mutual Fund Industry ", *Journal of Finance*, 54, 791-816.

Kosowski, R., (2011), "Do mutual funds perform when it matters most to investors? U.S. mutual fund performance and risk in recessions and expansions" *Quarterly Journal of Finance*, 1, 607-664.

Kosowski, R., Timmermann, A., Wermers, R., White, H., (2006), "Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis" *Journal of Finance*, 61, 2551-2595.

Lo, A., MacKinlay, A. C., (1990), "An econometric analysis of nonsynchronous trading" *Journal of Econometrics*, 45, 181-212.

Mackinlay, A. C., (1995), "Multifactor models do not explain deviations from the CAPM", *Journal of Financial Economics*, 38, 3-28.

Martin, I. W. R., (2013), "Consumption-based asset pricing with higher cumulants", *Review of Economics Studies*, 80, 745-773.

Moskowitz, T., (2000), "Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses: Discussion" *Journal of Finance*, 55, 1695-1703.

Newey, W. K., West, K. D., (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703-708.

Pyo, U., (2011), "Minimax price bounds in incomplete markets", *Journal of Economics and Finance*, 35, 274-295.

Ross, S. A., (1976), "The arbitrage theory of capital asset pricing", *Journal of Economics Theory*, 13, 341-360.

Ross, S. A., (2005), "Neoclassical finance", Princeton University Press, Princeton, NJ.

Sharpe, W. F., (1966), "Mutual fund performance", *Journal of Business*, 1, 2, 119-138.

Snow, K. N., (1991), "Diagnosing asset pricing models using the distribution of asset returns", *Journal of Finance*, 46, 955-983.

Table 2.1: Summary Statistics

Table 2.1 presents summary statistics for the monthly data from January 1984 to December 2012. Panel A shows cross-sectional summary statistics (average (Mean), standard deviation (StdDev) and selected percentiles) on the distributions of the average (Mean), standard deviation (StdDev), minimum (Min), maximum (Max), Sharpe ratio (h) and Carhart SDF alphas with their corresponding t-statistics for the returns on 2786 actively managed open-ended U.S. equity mutual funds. It also reports the t-statistics (t-stat) on the significance of the mean of estimated Carhart SDF alphas (see test description in section 2.3.3). Panel B gives the average (Mean), standard deviation (StdDev), minimum (Min), maximum (Max) and Sharpe ratio (h) for passive portfolio returns and information variables. Passive portfolios include ten industry portfolios (consumer nondurables (NoDur), consumer durables (Dur), manufacturing (Manuf), energy (Enrgy), high technology (HiTec), telecommunication (Telcm), shops (Shops), healthcare (Hlth), utilities (Utils), and other industries (Other)), six Fama-French style portfolios based on two market equity capitalization (size) sorts (big (B) or small (S)) and three book-to-market (value) sorts (low (L), medium (M) or high (H)), the market portfolio (MKT) based on the CRSP value-weighted index, and the risk-free asset (RF) based on the one-month Treasury bill. The information variables are lagged values of the dividend yield on the S&P500 Index (DIV), the yield on the three-month Treasury bill (YLD), the term spread (TERM) and the default spread (DEF). All statistics are in percentage except for the Sharpe ratios and the t-statistics.

Panel A: Mutual Fund Returns and Carhart Alphas							
	Mutual Fund Returns					Carhart Alphas	
	Mean	StdDev	Min	Max	h	α_{MF}	t-statistics
Mean	0.7338	5.3400	-19.9976	16.4080	0.0857	-0.1200	-0.7398
StdDev	0.3008	1.5632	5.6560	7.8772	0.0532	0.2456	1.2534
(t-stat)						(-2.321)	
99%	1.3677	10.3594	-5.5453	41.5517	0.1990	0.4231	2.2289
95%	1.1405	8.2353	-12.8253	32.5704	0.1593	0.2120	1.3169
90%	1.0453	7.2002	-14.4430	27.0565	0.1427	0.1271	0.8321
75%	0.9044	6.0719	-16.5652	18.5645	0.1178	0.0109	0.0731
Median	0.7464	5.0272	-19.4020	14.1103	0.0900	-0.1016	-0.7466
25%	0.5946	4.3865	-22.9302	11.4669	0.0613	-0.2214	-1.5149
10%	0.4232	3.8968	-26.3190	9.9918	0.0229	-0.3811	-2.2580
5%	0.2943	3.4811	-29.0886	9.0829	-0.0033	-0.5037	-2.7996
1%	-0.1139	1.6197	-36.9313	5.3715	-0.0827	-0.8329	-4.1538

Table 2.1: Summary Statistics (Continued)

Panel B: Passive Portfolio Returns and Information Variables					
	Mean	StdDev	Min	Max	*h*
Industry Portfolios					
NoDur	1.1713	4.3493	-21.0300	14.7400	0.1962
Durbl	0.8311	7.0347	-32.8900	42.9200	0.0698
Manuf	1.0667	5.1203	-27.3200	17.7800	0.1420
Enrgy	1.1223	5.3691	-18.3900	19.1300	0.1459
HiTec	0.9338	7.2260	-26.1500	20.4600	0.0822
Telcm	0.9689	5.2612	-15.5600	22.1200	0.1199
Shops	1.0394	5.0898	-28.3100	13.3800	0.1375
Hlth	1.1140	4.7552	-20.4700	16.5400	0.1636
Utils	0.9444	3.9952	-12.6500	11.7600	0.1521
Other	0.8829	5.3165	-23.6800	16.1100	0.1024
Style Portfolios, Market Portfolio and Risk-Free Asset					
B/L	0.9403	4.7004	-23.1900	14.4500	0.1281
B/M	0.9705	4.5832	-20.3200	14.8500	0.1378
B/H	0.9279	5.2367	-24.4700	22.1600	0.1126
S/L	0.8037	6.7657	-32.3400	27.0200	0.0685
S/M	1.1221	5.2548	-27.5700	18.8700	0.1487
S/H	1.2282	6.2180	-28.0500	38.3900	0.1426
MKT	0.9174	4.5814	-22.5363	12.8496	0.1262
RF	0.3393	0.2166	0.0000	1.0000	-
Information Variables					
DIV	2.4649	0.9204	1.0800	4.9900	-
YLD	4.1264	2.6038	0.0100	10.4700	-
TERM	1.9419	1.1392	-0.5300	3.7600	-
DEF	1.0255	0.4046	0.5500	3.3800	-

Table 2.2: Spanning Regression R-Squared, Disagreement Parameter Estimates and Stochastic Discount Factors for Performance Evaluation

Table 2.2 shows statistics on the cross-sectional distributions of R-squared values from the regressions of mutual fund returns on passive portfolio returns (panel A), estimates of the disagreement parameter v for two best clientele performance measures (denoted by $h^* + 0.5hMKT$ and $h^* + hMKT$; panel B), and estimates of the stochastic discount factors for two best clientele performance measures and for the LOP measure (denoted by h^*; panel C). The passive portfolios in panel A are the risk-free rate and either ten industry portfolios, six style portfolios, the market portfolio or the Carhart factors. The best clientele measures in panels B and C consider maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and all measures in panels B and C use the risk-free rate and ten industry portfolios (RF + 10I) as basis assets. Panel A provides the mean, standard deviation (StdDev), and 99th and 1st percentiles. Panel B provides the mean, standard deviation (StdDev), and proportions in percentage of estimated v that are significantly positive at the 5% ($\%v\ signif > 0\ at\ 5\%$) and 10% ($\%v\ signif > 0\ at\ 10\%$) levels. Panel C provides the mean, standard deviation (StdDev), 99th and 1st percentiles, and proportions in percentage of stochastic discount factor estimates that are positive ($\%m > 0$) and negative ($\%m > 0$). The data (see description in Table 2.1) cover the period January 1984-December 2012.

Panel A: R-squared from Regressions of Mutual Fund Returns on Passive Portfolio Returns

	Ten Industry Portfolios	Six Style Portfolios	Market Portfolio	Carhart factors
Mean	0.8157	0.8526	0.7475	0.8352
StdDev	0.1432	0.1436	0.1652	0.1450
99%	0.9723	0.9773	0.9650	0.9744
1%	0.1741	0.1391	0.0570	0.1152

Panel B: Estimates of Disagreement Parameter v Using the RF + 10I Passive Portfolio Set

	V		t-statistics	
	$h^* + 0.5hMKT$	$h^* + hMKT$	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	11.0036	13.7640	1.8285	3.9927
StdDev	5.5602	7.6070	0.3309	0.7389
$\%v\ signif > 0\ at\ 5\%$	28.07	100.00		
$\%v\ signif > 0\ at\ 10\%$	67.37	100.00		

Panel C: Stochastic Discount Factors Using the RF + 10I Passive Portfolio Set

	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	0.9971	0.9971	0.9971
StdDev	0.2571	0.3186	0.3819
99%	1.6621	1.8281	1.9967
1%	0.4538	0.2614	0.0777
$\%m > 0$	99.98	99.81	99.39
$\%m < 0$	0.02	0.19	0.61

Table 2.3: Best Clientele Alphas Using the RF + 10I Passive Portfolio Set

Table 2.3 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and with the LOP measure (denoted by h^*), using the risk-free rate and ten industry portfolios (RF + 10I) as basis assets. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). Panel B gives proportions of estimated alphas that are positive ($\%\bar{\alpha}_{MF} > 0$), negative ($\%\bar{\alpha}_{MF} < 0$), significantly positive ($\%\bar{\alpha}_{MF}signif > 0$), and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$). It also provides proportions adjusted for false discoveries according to the BSW and FC classifications (see description in section 2.3.3), i.e., proportions of zero alpha, unskilled and skilled funds. It finally presents the p-values (in parentheses) for the likelihood ratio tests (see description in section 2.3.3) that the proportions of positive estimated alphas are equal to 50%, and the proportions of significantly positive and significantly negative estimated alphas are equal to 2.5%. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics of Individual Mutual Funds

	Performance			t-statistics		
	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	-0.1789	0.2360	0.4436	-1.2291	1.0220	1.9033
StdDev	0.2707	0.3344	0.4183	1.4743	1.4333	1.4009
(t-stat)	(-3.138)	(3.351)	(5.036)			
99%	0.3580	1.2708	1.8383	2.0211	4.4955	5.4554
95%	0.1787	0.7851	1.1586	1.0473	3.3544	4.2584
90%	0.0967	0.6440	0.9764	0.5435	2.7535	3.5898
75%	-0.0364	0.4225	0.6801	-0.2258	1.9417	2.7639
Median	-0.1630	0.1841	0.3575	-1.1398	1.0652	1.8940
25%	-0.2854	0.0223	0.1494	-2.1715	0.1720	1.0671
10%	-0.4501	-0.1029	0.0182	-3.1627	-0.8232	0.1570
5%	-0.5978	-0.1909	-0.0708	-3.6814	-1.4254	-0.4560
1%	-0.9247	-0.4353	-0.2388	-4.8763	-2.6052	-1.4975

Panel B: Performance Proportions

		h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Performance	$\%\bar{\alpha}_{MF} > 0$	20.32 (0.00)	78.36 (0.00)	91.49 (0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	79.68	24.41	8.51
Performance	$\%\bar{\alpha}_{MF}signif > 0$	1.04 (0.00)	21.64 (0.00)	47.52 (0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	29.54 (0.00)	2.26 (40.96)	0.47 (0.00)
BSW Classification	Zero alpha	49.49	49.19	22.06
Adjusted for	Unskilled	50.51	0	0
False Discoveries	Skilled	0.00	50.81	77.94
FC Classification	Zero alpha	67.16	64.77	34.66
Adjusted for	Unskilled	32.84	0.00	0.00
False Discoveries	Skilled	0.00	35.23	65.34

Table 2.4: Conditional Best Clientele Alphas Using the RF + 10I + RZ Passive Portfolio Set

Table 2.4 shows statistics on the cross-sectional distribution of monthly conditional SDF alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and with the LOP measure (denoted by h^*), using the risk-free rate, ten industry portfolios and public information-managed payoffs (RF + 10I + RZ) as basis assets. Results are shown for average conditional alphas, as well as average conditional alphas in recessions and expansion. Panel A provides the mean, standard deviation (StdDev) and median of the distributions of estimated alphas. It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). Panel B gives proportions of estimated alphas that are positive ($\%\bar{\alpha}_{MF} > 0$), negative ($\%\bar{\alpha}_{MF} < 0$), significantly positive ($\%\bar{\alpha}_{MF}signif > 0$), and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$). It also provides proportions adjusted for false discoveries according to the BSW and FC classifications (see description in section 2.3.3), i.e., proportions of zero alpha, unskilled and skilled funds. It finally presents the p-values (in parentheses) for the likelihood ratio tests (see description in section 2.3.3) that the proportions of positive estimated alphas are equal to 50%, and the proportions of significantly positive and significantly negative estimated alphas are equal to 2.5%. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics of Individual Mutual Funds

	Performance			t-statistics		
	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Average Conditional Alphas						
Mean	-0.1795	0.2298	0.4347	-1.2283	1.0138	1.8980
Std Dev	0.2684	0.3322	0.4145	1.4582	1.4399	1.4186
(t-stat)	(-3.176)	(3.284)	(4.980)			
Median	-0.1642	0.1813	0.3524	-1.1368	1.0528	1.8940
Average Conditional Alphas in Expansions						
Mean	-0.1888	0.2200	0.4245	-1.4069	1.4647	2.6376
Std Dev	0.2576	0.3287	0.4134	1.5131	2.5582	3.1596
(t-stat)	(-3.481)	(3.178)	(4.877)			
Median	-0.1812	0.1682	0.3384	-1.2477	0.8822	1.7058
Average Conditional Alphas in Recessions						
Mean	-0.2005	0.2596	0.4884	-0.4067	1.7604	3.0569
Std Dev	1.1634	1.0654	1.1371	6.3434	7.7998	7.8361
(t-stat)	(-0.818)	(1.157)	(2.040)			
Median	0.1232	0.3999	0.5584	0.9875	2.0451	2.6832

Table 2.4: Conditional Best Clientele Alphas Using the RF + 10I + RZ Passive Portfolio Set (Continued)

Panel B: Performance Proportions				
		h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Average Conditional Alphas				
Performance	$\%\bar{\alpha}_{MF} > 0$	19.81(0.00)	77.93 (0.00)	91.03 (0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	80.19	22.07	8.97
Performance	$\%\bar{\alpha}_{MF}signif > 0$	1.01 (0.00)	24.23 (0.00)	47.67 (0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	29.40 (0.00)	2.15 (23.05)	0.36 (0.00)
BSW Classification	Zero alpha	49.06	49.53	22.62
Adjusted for	Unskilled	50.94	0.00	0.00
False Discoveries	Skilled	0.00	50.47	77.38
FC Classification	Zero alpha	67.64	65.63	34.95
Adjusted for	Unskilled	32.36	0.00	0.00
False Discoveries	Skilled	0.00	34.37	65.05
Average Conditional Alphas in Expansions				
Performance	$\%\bar{\alpha}_{MF} > 0$	16.12(0.00)	75.09(0.00)	89.91(0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	83.88	24.91	10.09
Performance	$\%\bar{\alpha}_{MF}signif > 0$	0.65(0.00)	30.29(0.00)	45.37(0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	34.06(0.00)	2.30(0.49)	0.75(0.00)
BSW Classification	Zero alpha	46.65	62.41	38.83
Adjusted for	Unskilled	53.35	0.00	0.00
False Discoveries	Skilled	0.00	37.59	61.17
FC Classification	Zero alpha	64.55	63.76	43.76
Adjusted for	Unskilled	35.45	0.00	0.00
False Discoveries	Skilled	0.00	36.24	56.24
Average Conditional Alphas in Recessions				
Performance	$\%\bar{\alpha}_{MF} > 0$	58.08(0.00)	75.41(0.00)	80.15(0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	41.92	24.59	19.85
Performance	$\%\bar{\alpha}_{MF}signif > 0$	38.05(0.00)	49.89(0.00)	56.96(0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	26.63(0.00)	15.51(0.00)	12.96(0.00)
BSW Classification	Zero alpha	22.21	26.15	20.43
Adjusted for	Unskilled	31.23	12.79	10.50
False Discoveries	Skilled	46.56	61.06	69.07
FC Classification	Zero alpha	33.46	27.70	21.12
Adjusted for	Unskilled	2.52	0.00	0.00
False Discoveries	Skilled	64.02	72.30	78.88

Table 2.5: Best Clientele Alphas Using the RF + 6S Passive Portfolio Set

Table 2.5 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and with the LOP measure (denoted by h^*), using the risk-free rate and six style portfolios (RF + 6S) as basis assets. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). Panel B gives the proportions of estimated alphas that are positive ($\%\bar{\alpha}_{MF} > 0$), negative ($\%\bar{\alpha}_{MF} < 0$), significantly positive ($\%\bar{\alpha}_{MF}signif > 0$), and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$). It also provides proportions adjusted for false discoveries according to the BSW and FC classifications (see description in section 2.3.3), i.e., proportions of zero alpha, unskilled and skilled funds. It finally presents the p-values (in parentheses) for the likelihood ratio tests (see description in section 2.3.3) that the proportions of positive estimated alphas are equal to 50%, and the proportions of significantly positive and significantly negative estimated alphas are equal to 2.5%. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics of Individual Mutual Funds

	Performance			t-statistics		
	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	-0.0839	0.2889	0.4724	-0.7286	1.5947	2.4692
StdDev	0.2680	0.3447	0.4168	1.5176	1.4563	1.4438
(t-stat)	(-1.487)	(3.980)	(5.383)			
99%	0.5424	1.4189	1.9912	2.7262	5.0331	6.0327
95%	0.3215	0.8744	1.1629	1.8065	3.9703	4.8819
90%	0.2063	0.7058	0.9627	1.2286	3.4759	4.3607
75%	0.0364	0.4394	0.6507	0.2899	2.5770	3.4205
Median	-0.0866	0.2174	0.3753	-0.7565	1.5263	2.4066
25%	-0.2015	0.0829	0.2119	-1.7059	0.6696	1.5284
10%	-0.3333	-0.0235	0.0948	-2.6708	-0.1833	0.7152
5%	-0.4451	-0.1086	0.0207	-3.2540	-0.8038	0.1600
1%	-0.7944	-0.3439	-0.1702	-4.3803	-1.8938	-0.9180

Panel B: Performance Proportions

		h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Performance	$\%\bar{\alpha}_{MF} > 0$	31.19 (0.00)	87.87 (0.00)	95.87 (0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	68.81	12.13	4.13
Performance	$\%\bar{\alpha}_{MF}signif > 0$	3.73 (0.01)	39.02 (0.00)	63.35 (0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	20.03 (0.00)	0.90 (0.00)	0.29 (0.00)
BSW Classification	Zero alpha	58.95	49.19	12.25
Adjusted for	Unskilled	38.28	0.00	0.00
False Discoveries	Skilled	2.78	50.81	87.75
FC Classification	Zero alpha	75.02	48.88	17.19
Adjusted for	Unskilled	24.98	0.00	0.00
False Discoveries	Skilled	0.00	51.12	82.81

Table 2.6: Best Clientele Alphas Using the RF + MKT Passive Portfolio Set

Table 2.6 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and with the LOP measure (denoted by h^*), using the risk-free rate and the market portfolio (RF + MKT) as basis assets. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). Panel B gives proportions of estimated alphas that are positive ($\%\bar{\alpha}_{MF} > 0$), negative ($\%\bar{\alpha}_{MF} < 0$), significantly positive ($\%\bar{\alpha}_{MF}signif > 0$), and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$). It also provides proportions adjusted for false discoveries according to the BSW and FC classifications (see description in section 2.3.3), i.e., proportions of zero alpha, unskilled and skilled funds. It finally presents the p-values (in parentheses) for the likelihood ratio tests (see description in section 2.3.3) that the proportions of positive estimated alphas are equal to 50%, and the proportions of significantly positive and significantly negative estimated alphas are equal to 2.5%. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics of Individual Mutual Funds

	Performance			t-statistics		
	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	-0.0683	0.2703	0.4703	-0.3830	1.1799	1.9011
StdDev	0.2747	0.3034	0.3655	1.1960	1.1825	1.1831
(t-stat)	(-1.180)	(4.231)	(6.110)			
99%	0.4663	1.1161	1.6469	2.2969	3.9490	4.8580
95%	0.2957	0.7570	1.0525	1.4196	3.0472	3.8673
90%	0.2171	0.6252	0.9001	1.0338	2.6149	3.4019
75%	0.0767	0.4427	0.6824	0.4046	1.9450	2.6431
Median	-0.0515	0.2531	0.4381	-0.3137	1.2177	1.8795
25%	-0.1845	0.0793	0.2136	-1.1147	0.4580	1.1640
10%	-0.3458	-0.0381	0.0829	-1.8898	-0.2371	0.5229
5%	-0.4818	-0.1437	-0.0015	-2.4349	-0.8009	-0.0157
1%	-0.8637	-0.4081	-0.2692	-3.7594	-1.9842	-1.1347

Panel B: Performance Proportions

		h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Performance	$\%\bar{\alpha}_{MF} > 0$	39.63 (0.00)	86.04 (0.00)	94.94 (0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	60.37	13.96	5.06
Performance	$\%\bar{\alpha}_{MF}signif > 0$	1.87 (2.50)	24.37 (0.00)	47.31 (0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	9.37 (0.00)	1.08 (0.00)	0.22 (0.00)
BSW Classification	Zero alpha	84.41	41.78	15.21
Adjusted for	Unskilled	15.59	0.00	0.00
False Discoveries	Skilled	0.00	58.22	84.79
FC Classification	Zero alpha	91.53	64.46	34.92
Adjusted for	Unskilled	8.47	0.00	0.00
False Discoveries	Skilled	0.00	35.54	65.08

Table 2.7: Best Clientele Alphas for Alternative Maximum Sharpe Ratio Choices Using the RF + 10I Passive Portfolio Set

Table 2.7 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with six best clientele performance measures, allowing for maximum Sharpe ratios of $1.5h^*$, $2h^*$, $h^* + 0.5hT$, $h^* + hT$, $h^* + 0.5hTa$ and $h^* + hTa$ (see definition in section 2.5.4), using the risk-free rate and ten industry portfolios (RF + 10I) as basis assets. It provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

	Performance and t-statistics of Individual Mutual Funds											
	Performance						t-statistics					
	$1.5h^*$	$2h^*$	$h^* + 0.5hT$	$h^* + hT$	$h^* + 0.5hTa$	$h^* + hTa$	$1.5h^*$	$2h^*$	$h^* + 0.5hT$	$h^* + hT$	$h^* + 0.5hTa$	$h^* + hTa$
Mean	0.4493	0.7973	0.4312	0.7669	0.2972	0.5428	1.9078	2.9954	1.8554	2.9386	1.3012	2.2612
StdDev	0.4197	0.5924	0.4127	0.5744	0.3571	0.4638	1.3900	1.3804	1.4022	1.3971	1.4220	1.3934
(t-stat)	(5.083)	(6.392)	(4.961)	(6.341)	(3.953)	(5.557)						
99%	1.8227	2.8781	1.8027	2.7665	1.4215	2.1207	5.3865	6.5569	5.3976	6.5673	4.8006	5.8459
95%	1.1659	1.8490	1.1321	1.7692	0.8968	1.3376	4.2270	5.4629	4.2018	5.4228	3.6333	4.6873
90%	0.9781	1.5523	0.9554	1.4971	0.7342	1.1374	3.5974	4.7727	3.5349	4.7671	3.0294	4.0023
75%	0.6834	1.1134	0.6655	1.0742	0.4995	0.8020	2.7638	3.8115	2.7182	3.7946	2.2002	3.0911
Median	0.3694	0.6661	0.3471	0.6416	0.2340	0.4455	1.9116	2.9080	1.8529	2.8461	1.3327	2.2237
25%	0.1518	0.3558	0.1417	0.3359	0.0592	0.2075	1.0726	2.1229	1.0197	2.0424	0.4554	1.4120
10%	0.0198	0.1967	0.0113	0.1896	-0.0668	0.0696	0.1690	1.3675	0.1077	1.2541	-0.5156	0.5470
5%	-0.0579	0.1241	-0.0793	0.1171	-0.1541	-0.0044	-0.4036	0.8194	-0.5104	0.7237	-1.1268	-0.0315
1%	-0.2559	-0.0336	-0.2503	-0.0190	-0.3882	-0.1543	-1.4611	-0.1925	-1.5536	-0.1024	-2.2417	-0.9806

Table 2.8: Bootstrap *p*-values and Proportions Using the RF + 10I Passive Portfolio Set

Table 2.8 shows statistics on the cross-sectional distribution of bootstrap *p*-values for alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and with the LOP measure (denoted by h^*), using the risk-free rate and ten industry portfolios (RF + 10I) as basis assets. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the bootstrap *p*-values. Panel B gives proportions of alphas that are significantly positive ($\%\bar{\alpha}_{MF}signif > 0$) and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$) using the bootstrap *p*-values. It also provides proportions adjusted for false discoveries, i.e., proportions of zero alpha, unskilled and skilled funds, based on the simulated critical values for the *t*-statistics that correspond to the size used for the BSW or FC classifications (see description in section 2.3.3). It finally presents the *p*-values (in parentheses) for the likelihood ratio tests (see description in section 2.3.3) that the proportions of significantly positive and significantly negative estimated alphas are equal to 2.5%. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage.

Panel A: Bootstrap Performance *p*-values			
	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	32.19	32.13	19.20
StdDev	31.43	30.88	26.23
99%	98.80	98.40	96.00
95%	91.60	91.80	80.80
90%	82.80	82.40	62.60
75%	56.80	56.60	29.60
Median	21.80	21.80	5.94
25%	3.20	4.20	0.40
10%	0.20	0.40	0.00
5%	0.00	0.00	0.00
1%	0.00	0.00	0.00

Panel B: Bootstrap Performance Proportions				
		h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Performance	$\%\bar{\alpha}_{MF}signif > 0$	0.93 (0.00)	22.47 (0.00)	46.73 (0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	28.46 (0.00)	4.31 (0.00)	1.08 (0.00)
BSW Classification Adjusted for False Discoveries	Zero alpha	52.18	58.22	27.40
	Unskilled	47.82	0.00	0.00
	Skilled	0.00	41.78	72.60
FC Classification Adjusted for False Discoveries	Zero alpha	69.19	68.71	38.47
	Unskilled	30.81	0.00	0.00
	Skilled	0.00	31.29	61.53

Table 2.9: Zero-Alpha Implied Maximum Sharpe Ratios

Table 2.9 shows statistics on the cross-sectional distribution of monthly maximum Sharpe ratios implied by fixing the best clientele alpha at zero (denoted by $\bar{\alpha}_{MF} = 0$), attainable monthly optimal Sharpe ratios of the passive portfolios (denoted by Basis Assets), and differences between both Sharpe ratios (denoted by Difference), using the risk-free rate and either ten industry portfolios, six style portfolios or the market portfolio as basis assets. It provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the values. The data (see description in Table 2.1) cover the period January 1984-December 2012.

	Sharpe Ratios								
	Ten Industry Portfolios			Six Style Portfolios			Market Portfolio		
	$\bar{\alpha}_{MF} = 0$	Basis Assets	Difference	$\bar{\alpha}_{MF} = 0$	Basis Assets	Difference	$\bar{\alpha}_{MF} = 0$	Basis Assets	Difference
Mean	0.2983	0.2571	0.0412	0.3087	0.2790	0.0296	0.1480	0.1146	0.0335
StdDev	0.0718	0.0408	0.0576	0.0556	0.0275	0.0449	0.0588	0.0383	0.0485
99%	0.5746	0.4628	0.2591	0.5384	0.3643	0.2186	0.3510	0.2713	0.2549
95%	0.4527	0.3002	0.1622	0.4119	0.3267	0.1135	0.2648	0.1626	0.1305
90%	0.3868	0.2838	0.1136	0.3676	0.3103	0.0763	0.2158	0.1484	0.0869
75%	0.3169	0.2692	0.0538	0.3223	0.2847	0.0373	0.1665	0.1270	0.0428
Median	0.2765	0.2528	0.0182	0.2924	0.2761	0.0134	0.1334	0.1171	0.0150
25%	0.2539	0.2364	0.0038	0.2777	0.2677	0.0033	0.1170	0.0965	0.0034
10%	0.2397	0.2214	0.0006	0.2674	0.2515	0.0005	0.0944	0.0700	0.0006
5%	0.2293	0.2147	0.0002	0.2540	0.2367	0.0001	0.0788	0.0593	0.0001
1%	0.2135	0.1895	0.0000	0.2356	0.2276	0.0000	0.0522	0.0100	0.0000

Table 2.10: Worst Clientele Alphas Using the RF + 10I Passive Portfolio Set

Table 2.10 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with two worst clientele performance measures, allowing for maximum Sharpe ratios of $h^* + 0.5hMKT$ and $h^* + hMKT$ (see definition in section 2.3.2), and with the LOP measure (denoted by h^*), using the risk-free rate and ten industry portfolios (RF + 10I) as basis assets. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). Panel B gives proportions of estimated alphas that are positive ($\%\bar{\alpha}_{MF} > 0$), negative ($\%\bar{\alpha}_{MF} < 0$), significantly positive ($\%\bar{\alpha}_{MF}signif > 0$), and significantly negative ($\%\bar{\alpha}_{MF}signif < 0$). It also provides proportions adjusted for false discoveries according to the BSW and FC classifications (see description in section 2.3.3), i.e., proportions of zero alpha, unskilled and skilled funds. It finally presents the p-values (in parentheses) for the likelihood ratio tests (see description in section 2.3.3) that the proportions of positive estimated alphas are equal to 50%, and the proportions of significantly positive and significantly negative estimated alphas are equal to 2.5%. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics of Individual Mutual Funds

	Performance			t-statistics		
	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$	h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Mean	-0.1789	-0.5938	-0.8014	-1.2291	-3.5688	-4.4825
StdDev	0.2707	0.3883	0.4796	1.4743	1.4261	1.4545
(t-stat)	(-3.138)	(-7.261)	(-7.936)			
99%	0.3580	-0.0667	-0.1978	2.0211	-0.5195	-1.5281
95%	0.1787	-0.1978	-0.3185	1.0473	-1.3712	-2.2725
90%	0.0967	-0.2572	-0.3861	0.5435	-1.8549	-2.7444
75%	-0.0364	-0.3625	-0.5015	-0.2258	-2.5970	-3.4673
Median	-0.1630	-0.5140	-0.6903	-1.1398	-3.4628	-4.3675
25%	-0.2854	-0.7159	-0.9670	-2.1715	-4.4925	-5.3940
10%	-0.4501	-1.0216	-1.3558	-3.1627	-5.3940	-6.3130
5%	-0.5978	-1.2511	-1.6315	-3.6814	-6.0399	-7.0541
1%	-0.9247	-1.8508	-2.3907	-4.8763	-7.2135	-8.2666

Panel B: Performance Proportions

		h^*	$h^* + 0.5hMKT$	$h^* + hMKT$
Performance	$\%\bar{\alpha}_{MF} > 0$	20.32 (0.00)	0.39 (0.00)	0.04 (0.00)
Sign	$\%\bar{\alpha}_{MF} < 0$	79.68	99.61	99.96
Performance	$\%\bar{\alpha}_{MF}signif > 0$	1.04 (0.00)	0.00 (0.00)	0.00 (0.00)
Significance	$\%\bar{\alpha}_{MF}signif < 0$	29.54 (0.00)	88.69 (0.00)	97.38 (0.00)
BSW Classification	Zero alpha	49.49	1.56	0.00
Adjusted for	Unskilled	50.51	98.44	99.96
False Discoveries	Skilled	0.00	0.00	0.04
FC Classification	Zero alpha	67.16	8.13	1.37
Adjusted for	Unskilled	32.84	91.87	98.63
False Discoveries	Skilled	0.00	0.00	0.00

Table 2.11: Total Performance Disagreement

Table 2.11 shows statistics on the cross-sectional distribution of monthly total performance disagreement from the best and worst clientele performance measures with a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 2.3.2), using the risk-free rate and either ten industry portfolios, six style portfolios or the market portfolio as basis assets. It provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated total disagreement values (columns under Value) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated disagreement values. Total performance disagreement is defined as the difference between best and worst clientele alphas. The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

| | Total Disagreement | | | | | |
| | Ten Industry Portfolios | | Six Style Portfolios | | Market Portfolio | |
	Value	t-statistics	Value	t-statistics	Value	t-statistics
Mean	0.8297	6.3779	0.7455	6.5374	0.6770	5.4709
StdDev	0.4818	1.8945	0.4177	1.7208	0.3732	1.8587
(t-stat)	(8.179)		(8.477)		(8.615)	
99%	2.6483	10.9180	2.7855	10.7928	2.2666	10.1666
95%	1.7199	9.6564	1.3909	9.4813	1.3370	8.8657
90%	1.4243	8.9265	1.1522	8.8512	1.1325	8.0961
75%	1.0413	7.6173	0.8576	7.5729	0.8104	6.7388
Median	0.7015	6.2729	0.6630	6.4852	0.5982	5.2337
25%	0.4761	5.0391	0.4985	5.3896	0.4384	4.0360
10%	0.3779	3.9845	0.3901	4.4984	0.3112	3.2524
5%	0.3387	3.3579	0.3458	3.8592	0.2657	2.8132
1%	0.2736	2.3733	0.2690	2.1814	0.1755	2.2428

Appendix 2.A: Additional Results on Alternative Maximum Sharpe Ratio Choices

Table 2.A1: Best Clientele Alphas for Alternative Maximum Sharpe Ratio Choices Using the RF + 6S Passive Portfolio Set

Table 2.A1 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with six best clientele performance measures, allowing for maximum Sharpe ratios of $1.5h^*$, $2h^*$, $h^* + 0.5hT$, $h^* + hT$, $h^* + 0.5hTa$ and $h^* + hTa$ (see definition in section 2.5.4), using the risk-free rate and six style portfolios (RF + 6S) as basis assets. It provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

Performance and t-statistics of Individual Mutual Funds

	Performance						t-statistics					
	$1.5h^*$	$2h^*$	$h^* + 0.5hT$	$h^* + hT$	$h^* + 0.5hTa$	$h^* + hTa$	$1.5h^*$	$2h^*$	$h^* + 0.5hT$	$h^* + hT$	$h^* + 0.5hTa$	$h^* + hTa$
Mean	0.5070	0.8336	0.5032	0.8272	0.4460	0.7300	2.6056	3.6689	2.5965	3.6593	2.3555	3.3887
StdDev	0.4290	0.5826	0.4300	0.5815	0.4057	0.5343	1.4315	1.4607	1.4440	1.4732	1.4441	1.4607
(t-stat)	(5.612)	(6.795)	(5.557)	(6.755)	(5.222)	(6.489)						
99%	2.0568	3.0482	2.0701	2.9757	1.9312	2.6753	6.1934	7.4169	6.1807	7.3761	5.9092	7.0363
95%	1.2594	1.8209	1.2278	1.7985	1.1265	1.6204	5.0098	6.1325	5.0009	6.1443	4.7509	5.8024
90%	1.0066	1.4777	1.0080	1.4722	0.9228	1.3332	4.4851	5.5998	4.4896	5.6057	4.2352	5.3185
75%	0.6886	1.0655	0.6868	1.0549	0.6203	0.9451	3.5186	4.5684	3.5481	4.5700	3.3115	4.3208
Median	0.4074	0.6936	0.4028	0.6901	0.3529	0.6033	2.5387	3.6057	2.5349	3.5986	2.2913	3.3433
25%	0.2371	0.4592	0.2327	0.4545	0.1928	0.3902	1.6615	2.6673	1.6476	2.6374	1.4268	2.3905
10%	0.1154	0.3101	0.1132	0.3034	0.0798	0.2483	0.8503	1.8516	0.8314	1.8336	0.6052	1.5958
5%	0.0509	0.2361	0.0394	0.2296	0.0053	0.1748	0.3139	1.4199	0.3067	1.3465	0.0305	1.1069
1%	-0.1142	0.0961	-0.1449	0.0646	-0.1918	0.0151	-0.7048	0.5640	-0.7545	0.4216	-1.0606	0.1196

Table 2.A2: Best Clientele Alphas for Alternative Maximum Sharpe Ratio Choices Using the RF + MKT Passive Portfolio Set

Table 2.A2 shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with two best clientele performance measures, allowing for maximum Sharpe ratios of $1.5h^*$ and $2h^*$ (see definition in section 2.5.4), using the risk-free rate and the market portfolio (RF + MKT) as basis assets. It provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 2.3.3). The data (see description in Table 2.1) cover the period January 1984-December 2012. All statistics are in percentage except the t-statistics.

	Performance and t-statistics of Individual Mutual Funds				
	Performance			t-statistics	
	$1.5h^*$	$2h^*$		$1.5h^*$	$2h^*$
Mean	0.2505	0.4352		1.1110	1.7837
StdDev	0.3049	0.3757		1.2105	1.2607
(t-stat)	(3.901)	(5.501)			
99%	1.1289	1.6323		3.9514	4.8521
95%	0.7287	1.0219		3.0206	3.8211
90%	0.5985	0.8716		2.5699	3.3744
75%	0.4220	0.6381		1.8873	2.5850
Median	0.2280	0.4002		1.1412	1.7706
25%	0.0601	0.1835		0.3614	1.0109
10%	-0.0588	0.0505		-0.3316	0.2718
5%	-0.1636	-0.0524		-0.8830	-0.3185
1%	-0.4768	-0.3821		-2.2526	-1.6330

Figure 2.1: Mean-Standard Deviation Frontiers from the Sets of Passive Portfolio Returns
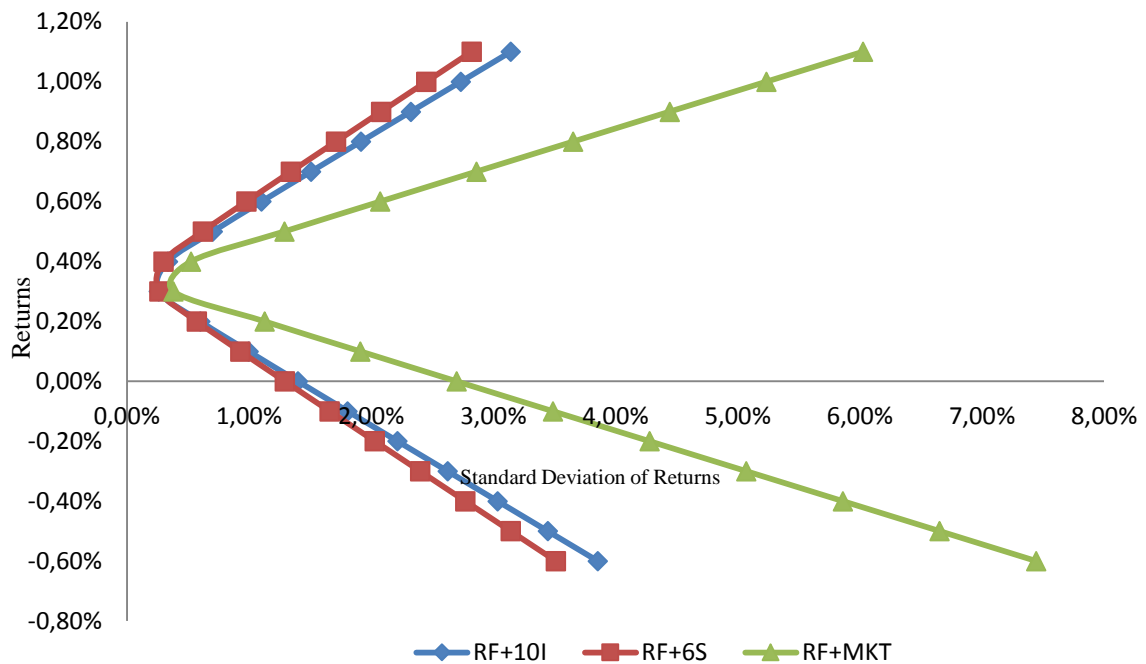
Notes: Figure 2.1 presents mean-standard deviation frontiers of investment opportunities from the risk-free rate and either ten industry portfolios (RF+10I), six style portfolios (RF+6S) or the market portfolio (RF+MKT) as passive portfolios.

Figure 2.2: Histograms of the Best Clientele and LOP Alphas
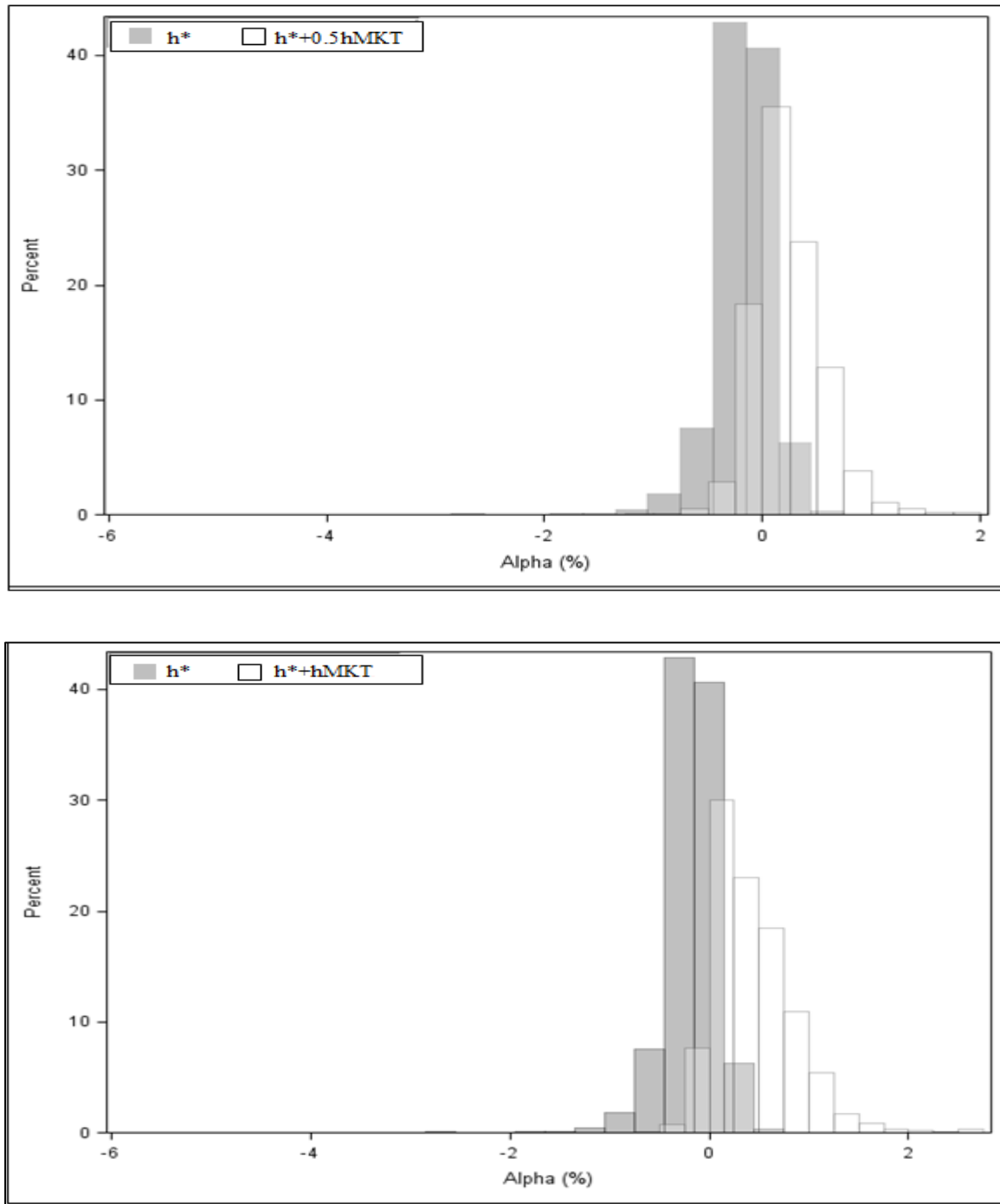
Notes: Figure 2.2 presents histograms illustrating the distributions of the best clientele and LOP alpha estimates, using the risk-free rate and ten industry portfolios as basis assets. Figure 2.2a illustrates LOP alphas (denoted by h^*) and best clientele alphas allowing for a maximum Sharpe ratio of $h^* + 0.5hMKT$. Figure 2.2b illustrates LOP alphas and best clientele alphas allowing for a maximum Sharpe ratio of $h^* + hMKT$.

Figure 2.3: Best Clientele and LOP Alphas for Decile Fund Portfolios

(a)

(b)

(c)
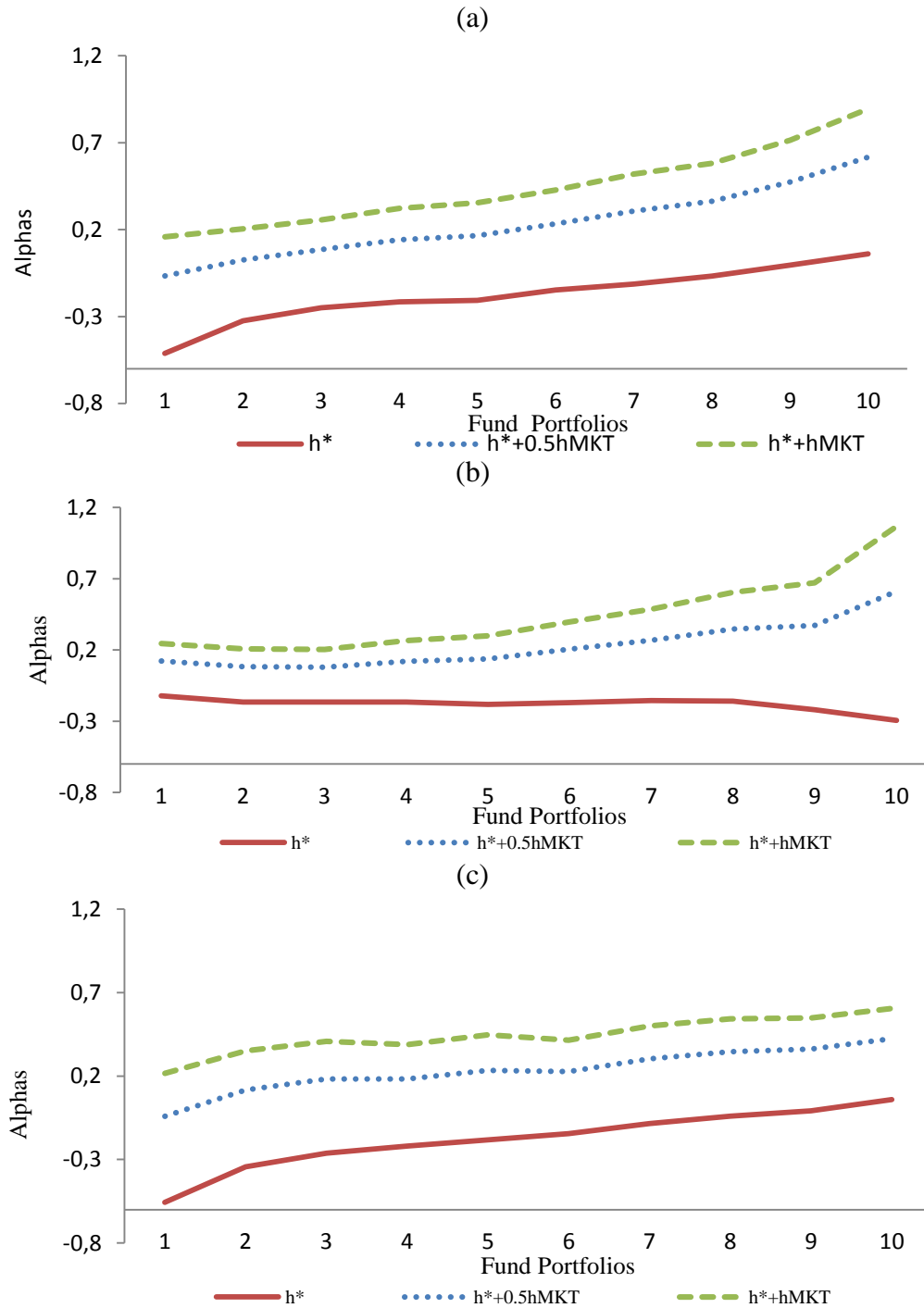
Notes: Figure 2.3 presents best clientele alpha estimates (denoted by $h^* + 0.5hMKT$ and $h^* + hMKT$) and LOP alpha estimates (denoted by h^*) for mutual funds grouped in decile portfolios. In Figure 2.3a, funds are sorted in increasing order of their average return. In Figure 2.3b, funds are sorted in increasing order of their standard deviation of returns. In Figure 2.3c, funds are sorted in increasing order of their Sharpe ratio.

3 Representative Investors versus Best Clienteles: Performance Evaluation Disagreement in Mutual Funds

Abstract

The paper examines performance disagreement in mutual funds and develops a diagnostic tool for candidate performance models. We compare the evaluation for best clienteles, as specified by an upper admissible performance bound, to the one for representative investors implicit in twelve models. Empirical results show that the linear factor models are the most severe in their assessment because they significantly undervalue the best clientele alphas. Consumption-based models provide alphas that tend to be inadmissible because they are too high. The manipulation proof performance measure generates alphas that are sensitive to the choice of risk aversion parameter and difficult to estimate with statistical precision. However, a reasonable risk aversion parameter generally gives admissible values that adequately reflect the alphas for the most favorable clienteles. Our results support an economically important role for performance evaluation disagreement in mutual funds.

Résumé

Cet article examine le désaccord entre investisseurs dans les fonds mutuels et développe un outil pour diagnostiquer des modèles de performance candidats. Nous comparons l'évaluation pour les meilleures clientèles, spécifiée par une borne supérieure de performance, à celle pour les investisseurs représentatifs implicites dans douze modèles. Les résultats empiriques montrent que les modèles linéaires à facteurs sont les plus sévères puisqu'ils sous-évaluent significativement les alphas des meilleures clientèles. Les modèles de consommation procurent des alphas qui ont tendance à être inadmissibles puisque trop élevés. La mesure de performance à l'abri de manipulation génère des alphas qui sont sensibles au choix du paramètre d'aversion au risque et difficiles à estimer avec précision statistique. Toutefois, un paramètre d'aversion au risque raisonnable donne généralement des alphas admissibles qui reflètent adéquatement les alphas des clientèles les plus favorables. Les résultats supportent un rôle économique important pour le désaccord entre investisseurs dans l'évaluation de performance des fonds mutuels.

84

3.1 Introduction

Disagreement is undoubtedly a part of life. In politics, voters often disagree about the best policies for their country and are attracted to various political options based on their beliefs. In sports, fans are passionate about their teams and often dislike their rival teams. At home, households disagree on how to raise children, what goods are needed for consumption, where to spend their leisure time, etc. In fact, in many instances, it is not uncommon for people to "agree to disagree". Similarly, in portfolio choice, investors clearly disagree on what investments are the most valuable for their portfolios.

Recently, Ferson and Lin (2014) demonstrate the large and significant effects of investor disagreement on the performance evaluation of equity mutual funds. In particular, they develop bounds on the expected disagreement with a traditional alpha and document average monthly values from 0.21% to 0.38%. Given these values, they argue that the effect of investor heterogeneity on alpha could be as important as the already well documented effects of the benchmark choice or statistical imprecision of estimates. Their study is part of a growing literature on studying specific types of investor heterogeneity and clienteles in the mutual fund sector (i.e., Nanda, Narayanan and Warther, 2000; Malloy and Zhu, 2004, Barber, Odean and Zheng, 2005, Bergstresser, Chalmers and Tufano, 2009; Nanda, Wang and Zheng, 2009, Bailey, Kumar and Ng, 2011; Christoffersen, Evans and Musto, 2013; Del Guercio and Reuter, 2014), and calls for more research on the issue.

In this paper, we examine performance evaluation disagreement in mutual funds with a new approach. We compare alphas from standard measures with an alpha from a measure that values a fund from the point of view of their most favorable investors. This latter evaluation, termed the "best clientele alpha" by Chrétien and Kammoun (2015), is obtained by estimating the upper admissible performance bound under the law-of-one-price and no-good-deal conditions, a framework initially developed by Cochrane and Saá-Requejo (2000) to bound option prices. This paper thus studies performance disagreement between two important groups of investors: *representative investors*, on whom standard measures typically rely for evaluation, versus *best clienteles*, who are potentially the most valuable targets for funds that cater to specific clienteles.

Apart from developing a new measure of alpha disagreement, our approach provides an interesting diagnostic tool for candidate performance models. When alphas for representative investors and best clienteles are not equal, two alternative hypotheses are insightful. On one hand, an inadmissibility problem occurs when a candidate alpha is greater than the upper admissible bound that is the best clientele alpha.[1] On the other hand, an issue we call the misrepresentation problem occurs when a candidate alpha is lower than the best clientele alpha, because the candidate alpha provides a "severe" or "pessimistic" evaluation of the fund that does not adequately reflect the more useful evaluation from its best potential clienteles. The misrepresentation problem is indicative of large investor disagreement in performance evaluation.

Given the abundance of performance measures, there is not only a need for evaluating the measures themselves, but our approach is also uniquely positioned compared with the simulation approach typically used in the literature. Kothari and Warner (2001), Farsnwoth, Ferson, Jackson and Todd (2002), Kosowski, Timmermann, Wermers and White (2006) and Chrétien, Coggins and d'Amours (2015) are examples of studies that rely on fictitious managers to assess the ability of various performance measures. In comparison to these studies, our diagnostic tool provides a different perspective on the appropriateness of performance measures that is more in line with the bound approach initiated by Hansen and Jagannathan (1991) for asset pricing models.

Empirically, we perform our investigation on twelve candidate models: four unconditional linear factor models (the CAPM and the models of Fama and French (1993), Carhart (1997) and Ferson and Schadt (1996)), four conditional linear factor models (conditional versions of the previous four models), two consumption-based models (a power utility model and an external habit-formation preference model first examined by Cochrane and Hansen (1992)), the manipulation proof performance measure (MPPM) of Goetzmann, Ingersoll, Spiegel and Welch (2007) and the nonparametric law-of-one-price (LOP) measure of Chen and Knez (1996). We use a sample of 2786 actively-managed

---

[1] See Chen and Knez (1996) for the minimum set of requirements for the admissibility of performance measures and Hansen and Jagannathan (1991) for a classic diagnostic tool on the admissibility of asset pricing models.

open-ended U.S. equity mutual funds with returns from 1984 to 2012. Our implementation of the MPPM is particularly noteworthy because it is the first to provide monthly effective MPPM alpha estimates, their standard errors and their t-statistics for a large sample of equity mutual funds. To our knowledge, no estimation strategy allowing for statistical inferences on the significance of the performance values exists in the MPPM literature.

The empirical results can be summarized as follows. The comparison between the candidate and best clientele alphas shows that many standard measures misrepresent the value of mutual funds for their most favorable clienteles. Specifically, the unconditional linear factor models, their conditional versions, the MPPM with high risk aversion parameter and the LOP measure tend to give a severe but admissible performance evaluation. The average performance disagreement between the best clienteles and the representative investors from these candidate models, measured by the mean difference between their alphas, vary from 0.283% to 0.464%. These statistically significant disagreement values are comparable to those in Ferson and Lin (2014) and reaffirm the economic importance of their results. Oppositely, the alphas from consumption-based models are oftentimes not admissible because they are too high. Among all models, the MPPM with a low risk aversion parameter is most appropriate in providing admissible alphas that reflect the value of funds for their best clienteles.

Our implementation of the MPPM documents three other interesting empirical findings. First, the MPPM alphas are relatively sensitive to the choice of risk aversion parameter. When the MPPM representative investors have low concerns about risk and manipulation, their alphas tend to be positive, but when they have high concerns, their alphas are the most negative of all models, consistent with traditional alphas being inflated by manipulation for investors who care about it. Second, the MPPM alphas are difficult to estimate with statistical precision and the proportions of funds with significant values are much lower. Third, the MPPM provides a larger standard deviation of alphas across funds compared with the other models because the alpha distributions present a large negative skewness. For a small number of funds, manipulation-proofing their performance gives significantly lower alphas than their traditional or best clientele alphas.

The remainder of this paper is organized as follows. Section 3.2 develops a theoretical framework for measuring performance disagreement and assessing the appropriateness of candidate measures, and describes the best clientele and standard performance measures. Section 3.3 presents the methodology for estimating and comparing alphas. Section 3.4 describes the data. Section 3.5 reports and analyzes the empirical results. Section 3.6 provides our conclusion.

3.2   Representative Investors versus Best Clienteles: A Yardstick for Comparison

This section develops a framework for comparing standard performance measures with a measure that accounts for the most favorable clienteles of a fund, and for assessing their disagreement. First, we provide a general setup within which the performance measures can be understood, and provide null and alternative hypotheses for their comparison. Second, we develop the best clientele performance measure that serves as a basis for comparison. Third, we present standard performance measures that are considered candidate models for diagnosis in this study.

3.2.1   A Framework for Performance Evaluation and Comparison

This paper uses the stochastic discount factor (hereafter SDF) performance evaluation approach first examined by Glosten and Jagannathan (1994) and Chen and Knez (1996). Ferson (2010) summarizes its benefits. In this approach, the SDF alpha is defined from the following equation:

(1) $$\alpha_{i,MF} = E[m_{it}\, R_{MFt}] - 1,$$

where m_{it} is the SDF at time t of investor i interested in valuing the mutual fund with gross return R_{MFt}, and $\alpha_{i,MF}$ is her average performance evaluation or alpha. In asset pricing, the SDF is the investor's intertemporal marginal rate of substitution. An intuitive interpretation of this equation is thus that the "marginal preferences" of investor i determine her alpha.

As preferences vary by investor, this equation leads to a performance evaluation that differs by investor. In fact, under general conditions with an incomplete market and potentially heterogeneous preferences, Chen and Knez (1996) show that there is an infinite number of SDFs, and argue that the alpha disagreement they generate might not be unrealistic. Empirically, Ferson and Lin (2014) and Chrétien and Kammoun (2015) find that such disagreement is economically important. Let M be the set of SDFs for all investors. This paper is interested in studying the performance disagreement between two important groups of investors within that set.

The first group, called *representative investors*, considers the SDF of "average" (or more precisely marginal) investors obtained under restrictive assumptions. By making assumptions on preferences, return distributions and/or market completeness, equilibrium or no arbitrage conditions lead to a unique parametric SDF that represents all investors and can be used for performance evaluation. We denote by $m_{\varphi t}(\boldsymbol{\theta})$ such a SDF and by $\alpha_{\varphi,MF}(\boldsymbol{\theta})$ its resulting alpha, where we explicitly acknowledge their parametric nature by making them dependent on a vector $\boldsymbol{\theta}$ of parameters. For example, the celebrated CAPM implies a representative investor with a SDF linear in the market portfolio return. Most studies focus exclusively on this group of investors for performance evaluation because they rely on parametric SDFs derived from commonly-used asset pricing models. The main advantage of these SDFs is that they provide unique point estimates for fund evaluation that are relevant for "average" investors.

The second group, called *best clienteles*, represents investors that value the fund the most favorably. This group is potentially the most important for mutual funds that, in practice, cater to different clienteles through their management style and other attributes, instead of simply marketing themselves to "average" investors. By making minimal assumptions on the set M, so that it includes only SDFs admissible under selected criteria, Chen and Knez (1996), Ahn, Cao and Chrétien (2009) and Chrétien and Kammoun (2015) show that there is an upper bound on the performance of a fund:

$$(2) \qquad \bar{\alpha}_{BC,MF} = \sup_{m_i \in M} E[m_{it} \, R_{MFt}] - 1.$$

The solution to this problem is a nonparametric performance evaluation measure termed the "best clientele measure" by Chrétien and Kammoun (2015), with its corresponding BC alpha denoted by $\bar{\alpha}_{BC,MF}$. It considers the performance from the point of view of investors who value the fund the most favorably, i.e. best potential clienteles of a fund. It also has the added benefit that it is admissible under the minimum set of requirements of Chen and Knez (1996), and thus represents an upper bound restriction that all performance measures should meet for admissibility.

Given these positive attributes, this paper uses the BC alpha $\bar{\alpha}_{BC,MF}$ as a relevant yardstick for comparison with a standard performance alpha $\alpha_{\varphi,MF}(\boldsymbol{\theta})$. As null hypothesis, we examine if a candidate performance alpha is equal to the BC alpha:

$$H_0: \alpha_{\varphi,MF}(\boldsymbol{\theta}) = \bar{\alpha}_{BC,MF}.$$

Under this hypothesis, the performance measure φ gives a performance value that is admissible and reflects adequately the value for the most favorable clienteles of the fund. Thus, the representative investors behind the parametric SDF φ give a performance value that corresponds to the one for best clienteles, and hence φ is appropriate for a mutual fund that targets such a group. In the case where H_0 does not hold, two alternative hypotheses are interesting to understand why the candidate model does not meet this appropriateness criteria.

First, the alpha from a candidate model is higher than the BC alpha:

$$H_{a1}: \alpha_{\varphi,MF}(\boldsymbol{\theta}) > \bar{\alpha}_{BC,MF}.$$

If a standard measure provides a performance value significantly higher than the value from the upper admissible bound that represents the BC alpha, then it is inadmissible. Most likely, as discussed in Chen and Knez (1996), it suffers from the benchmark choice problem in that it is unable to correctly price passive portfolios. The related literature

90

(Lehmann and Modest, 1987, Grinblatt and Titman, 1994, Ahn, Cao and Chrétien, 2009, Cremers, Petajisto and Zitzewitz, 2013, among others) shows that choosing an inefficient benchmark may lead to a biased performance evaluation. The alternative hypothesis H_{a1} is helpful in detecting when this bias is positive, possibly indicating that the standard measure overvalues passive portfolios.

Hence, the BC alpha $\bar{\alpha}_{BC,MF}$ allows a partial diagnosis of such a problem among candidate measures, because a positive bias could result in a candidate alpha being above the upper bound. This analysis is similar in purpose to the investigation of Hansen and Jagannathan (1991), who use a SDF volatility bound to diagnose candidate models in the context of asset pricing. It complements the work of Ahn, Cao and Chrétien (2009), who propose a conservative check for the benchmark choice problem by comparing a candidate alpha with no arbitrage performance bounds and find that such bounds are oftentimes too wide to be a useful diagnostic tool.

Second, the alpha from a candidate model is lower than the BC alpha:

$$H_{a2}: \alpha_{\varphi,MF}(\boldsymbol{\theta}) < \bar{\alpha}_{BC,MF}.$$

If a standard measure gives a value that is significantly lower than the value for the best clienteles of the fund, then it is admissible but indicates that there is a large disagreement in fund evaluation. It implies that the representative investors assumed behind the standard models give a "severe" or "pessimistic" fund evaluation, which differs from the more useful evaluation that the best potential clienteles would give. We call such an outcome the "misrepresentation problem". The alternative hypothesis H_{a2} is thus helpful in detecting if the misrepresentation problem is important for candidate measures.[2]

By not considering investor disagreement in performance evaluation, the standard measures are inappropriate in considering the large set of investors who might be interested

---

[2] As noted by Ahn, Cao and Chrétien (2009), the alpha of a candidate model should also be greater than the lower admissible performance bound. It is thus possible that a measure suffers so badly from the misrepresentation problem as to render it inadmissible with respect to the lower bound. In section 3.5.7, we examine whether this is the case for the candidate models that are empirically found to have the most negative performance values.

in funds, including their most favorable clienteles. The literature has recently established the significance of heterogeneous preferences in mutual fund investors, so that they may look differently at the attractiveness of the same fund (see Ahn, Cao and Chrétien, 2009, Bailey, Kumar and Ng, 2011, Del Guercio and Reuter, 2013, Ferson and Lin, 2014, and Chrétien and Kammoun, 2015). The upper admissible performance bound $\bar{\alpha}_{BC,MF}$ allows a new analysis of the importance of investor disagreement in mutual fund evaluation and the related misrepresentation problem in standard performance models that ignore such a possibility.

3.2.2 Best Clientele Performance Measure

One key to the analysis of the previous subsection is the BC performance alpha that serves as a basis for comparison. Following Cochrane and Saá-Requejo (2000) and Chrétien and Kammoun (2015), this subsection presents the development and solution of the BC performance measure. These references should be consulted for a more complete analysis.

The first step to obtain $\bar{\alpha}_{BC,MF}$ is to impose a relevant structure on the set of SDFs of all investors to achieve a restricted set useful to identify the most favorable clienteles. We rely on the literature on asset pricing in incomplete markets to impose two restrictions. The first restriction is the *law-of-one-price condition* (Hansen and Jagannathan, 1991), which assumes that mutual fund investors give zero performance to passive portfolios. This restriction is equivalent to assuming a minimum variability in SDFs of investors. As explained by Chen and Knez (1996), Ahn, Cao and Chrétien (2009) and Chrétien and Kammoun (2015), imposing this condition can alleviate the benchmark choice problem. The second restriction is the *no-good-deal condition*, proposed by Cochrane and Saá-Requejo (2000). This condition assumes that investors do not allow for investment opportunities that have too large Sharpe ratios. Such opportunities are deemed too good to be viable because investors would quickly profit from them until they disappear. This restriction is equivalent to assuming a maximum variability in the SDFs of investors to preclude them from allowing implausibly good deals to exist.

As shown by Cochrane and Saá-Requejo (2000), imposing these two restrictions provides an upper bound by solving the following problem:

92

$$(3) \qquad \bar{\alpha}_{BC,MF} = \sup_{m_i \in M} E[m_{it} R_{MFt}] - 1,$$

$$(4) \qquad \text{subject to } E[m_{it} \mathbf{R_{Kt}}] = \mathbf{1}, E[m_{it}^2] \leq \frac{(1+\bar{h}^2)}{R_F{}^2},$$

where $E[m_{it} \mathbf{R_{Kt}}] = \mathbf{1}$ is law-of-one-price condition, with $\mathbf{R_{Kt}}$ being a vector of gross returns on K passive portfolios at time t and $\mathbf{1}$ is a $K \times 1$ unit vector, and $E[m_{it}^2] \leq \frac{(1+\bar{h}^2)}{R_F{}^2}$ is the no-good-deal condition, with $\bar{h}$ being the maximum Sharpe ratio allowed and R_F being the gross risk-free rate. Cochrane and Saá-Requejo (2000) demonstrate that this problem has the following solution:

$$(5) \qquad \bar{\alpha}_{BC,MF} = E[\bar{m}_{BCt} R_{MFt}] - 1,$$

with:

$$(6) \qquad \bar{m}_{BCt} = m_{LOPt} + v w_t,$$

$$(7) \qquad m_{LOPt} = \mathbf{a}' \mathbf{R_{Kt}},$$

$$(8) \qquad w_t = R_{MFt} - \mathbf{c}' \mathbf{R_{Kt}},$$

where:

$$(9) \qquad \mathbf{a}' = \mathbf{1}' E[\mathbf{R_{Kt}} \mathbf{R_{Kt}'}]^{-1},$$

$$(10) \qquad \mathbf{c}' = E[R_{MFt} \, \mathbf{R'_{Kt}}] \, E[\mathbf{R_{Kt}} \, \mathbf{R'_{Kt}}]^{-1},$$

$$(11) \qquad v = \sqrt{\frac{\left(\dfrac{(1 + \bar{h}^2)}{R_F{}^2} - E[m_t^{*2}]\right)}{E[w_t^2]}}.$$

In this solution, $\bar{m}_{BCt}$ represents the BC SDF and has two parts. The first part is m_{LOPt}, the SDF identified by Hansen and Jagannathan (1991) as having minimum volatility under the law-of-one-price condition and used by Chen and Knez (1996) for their LOP performance measure. Its volatility corresponds to $\sigma[m_{LOPt}] = h^*/R_F$, where h^* is the optimal Sharpe ratio obtained from passive portfolio returns. The second part is $v w_t$. The replicating error term w_t represents the difference between the mutual fund return and the best replicating payoff that can be constructed from passive portfolio returns. The parameter v accounts for the no-good-deal restriction and is a function of the maximum Sharpe ratio $\bar{h}$.

As discussed by Cochrane and Saá-Requejo (2000) and Chrétien and Kammoun (2015), there are two sources of investor disagreement within the BC performance measure. First, disagreement is increasing in the fund replicating error, so that investors have higher disagreement for funds with returns more difficult to span. Second, disagreement is increasing in the maximum Sharpe ratio, so that investors have higher disagreement when they consider more opportunities as "reasonable" (not being good deals). Finally, two other elements about the BC measure are noteworthy. First, it is an admissible measure in the sense that it respects the minimum set of requirements established by Chen and Knez (1996) for the admissibility of a performance measure (see also Hansen and Jagannathan (1991) on the related admissibility of SDFs for asset pricing models). The most important requirement is to respect the aforementioned law-of-one-price condition. Second, being an upper performance bound, it can be interpreted as the performance from the class of investors most favorable to a mutual fund. It is thus possible to evaluate whether a fund adds value from the point of view of their best potential clienteles. In particular, Chen and

Knez (1996) and Ferson and Lin (2014) show that if this value is positive, there are some investors that would want to buy the fund, with an optimal investment proportional to the alpha.

3.2.3 Candidate Performance Evaluation Measures

This subsection presents the menu of candidate performance measures that we compare with the BC measure to analyze their admissibility and disagreement. Although the BC measure is valid for evaluating portfolios with any type of assets, we focus on equity models because our empirical sample consists of equity mutual funds. We consider various classes of models, including linear factor models, conditional models, consumption-based models, a manipulation-proof measures and a nonparametric approach.

3.2.3.1 Unconditional Linear Factor Models

Return-based linear factor models are the most widely used models in performance evaluation. Their SDF is simply a linear function of factors:

$$(12) \qquad\qquad m_{\varphi t}(\boldsymbol{\theta}) = \omega_0 + \boldsymbol{\omega_1'}\mathbf{f_t},$$

where $\mathbf{f_t}$ is a vector of factors at time t.

The most well-known example of linear factor models is the classic CAPM, which implies that the SDF is a linear function of the market portfolio return such that:

$$(13) \qquad\qquad m_{CAPMt}(\boldsymbol{\theta}) = a_0 + a_{MKT}MKT_t,$$

where MKT_t is the market factor at time t and $\boldsymbol{\theta} = \{a_0, a_{MKT}\}$ are the parameters. Despite its popularity, the CAPM is widely criticized. In particular, Fama and French (1993) introduce two factors to better capture the variation of excess returns: the return difference between small and big market capitalization portfolios (SMB) and the return difference between high and low book-to-market ratio portfolios (HML). To further improve the

Fama-French model, Carhart (1997) includes a momentum factor (MOM), defined as the return difference between the winner and loser portfolios based on the prior one-year return. The SDFs implied by these models are expressed, respectively, as follows:

(14) $\quad m_{FFt}(\boldsymbol{\theta}) = b_0 + b_{MKT}(MKT_t - r_{ft}) + b_{SMB}SMB_t + b_{HML}HML_t,$

(15) $\quad m_{CARHARTt}(\boldsymbol{\theta}) = c_0 + c_{MKT}(MKT_t - r_{ft}) + c_{SMB}SMB_t + c_{HML}HML_t +$
$\quad c_{MOM}MOM_t,$

where r_{ft} is the net risk free return. Ferson and Schadt (1996) propose other factors to evaluate mutual fund performance: returns on large stocks (LS), small stocks (SS), long-term government bonds (LTGB) and low-grade corporate bonds (LGCB), with a resulting SDF given by:

(16) $\quad m_{FSt}(\boldsymbol{\theta}) = d_0 + d_{LS}LS_t + d_{SS}SS_t + d_{LTGB}LTGB_t + d_{LGCB}LGCB_t.$

Although many other linear factor models exist, these models are the most common in performance evaluation of equity portfolios.

3.2.3.2 Conditional Linear Factor Models

To account for time-varying economic conditions in financial markets, Ferson and Schadt (1996) advocate conditional versions of linear factor models that assume that factor coefficients are linear functions of publicly available information variables. The conditional SDF of a linear factor model has the following general form:

(17) $\quad m_{C\varphi t}(\boldsymbol{\theta}) = \beta_0 + \boldsymbol{\beta_1}(\mathbf{Z_{t-1}})'\mathbf{f_t},$

where $\boldsymbol{\beta_1}(\mathbf{Z_{t-1}}) = \boldsymbol{\beta_{10}} + \boldsymbol{\beta_{11}}'\mathbf{Z_{t-1}}$ and $\mathbf{Z_{t-1}}$ is a vector of information variables available at the time of investing. This paper considers a conditional version of the four linear factor

models presented earlier, leading to models denoted as CCAPM, CFF, CCARHART and CFS.

3.2.3.3 Consumption-based models

The SDF of consumption-based models is a function of the marginal utility of the representative agent. Although these fundamental models have been widely used in the equity premium puzzle literature, they are uncommon in the performance evaluation literature, perhaps because consumption data are problematic. We consider two consumption-based models. The first model assumes that the representative agent has time-separable power utility (POWER). The second model assumes an external habit-formation preference specification (HABIT) similar to the one proposed by Abel (1990) and Cochrane and Hansen (1992). The real SDFs implied by the two models are adjusted for inflation to obtain the following nominal versions:

(18)
$$m_{POWERt}(\boldsymbol{\theta}) = \beta \left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} \frac{PL_{t-1}}{PL_t},$$

(19)
$$m_{HABITt}(\boldsymbol{\theta}) = \beta \left(\frac{C_t - \mu C_{t-1}}{C_{t-1} - \mu C_{t-2}}\right)^{-\gamma} \frac{PL_{t-1}}{PL_t}.$$

where C_t is the consumption of the representative agent and PL_t is the price level at time t. The parameters are the subjective time discount factor β, the curvature parameter γ (which represents the relative risk aversion in the POWER model) and the habit parameter $\mu < 1$, which accounts for the fraction of previous consumption that the agent has become used to.

3.2.3.4 Manipulation-Proof Performance Measure

The manipulation-proof performance measure (MPPM), developed by Goetzmann, Ingersoll, Spiegel and Welch (2007), overcomes the problem that managers can manipulate existing performance measures. It is similar to a SDF alpha for a power utility function of the market portfolio return, with a relative risk aversion set so that holding the market

portfolio is optimal. The specific measure proposed by the authors provides an annualized continuously compounded excess return certainty equivalent for a mutual fund that can be calculated as follows:

$$(20) \qquad MPPM = \frac{1}{(1-\gamma)\Delta t} \ln \left(\frac{1}{T} \sum_{t=1}^{T} \left[\frac{R_{MFt}}{R_{Ft}} \right]^{1-\gamma} \right),$$

where T is the total number of observations, γ is the risk aversion, and Δt is the length of time in years between observations. Because the SDF alphas from other models are monthly effective excess return certainty equivalent, we transform MPPM to a value in comparable unit by using the following equivalence:

$$(21) \qquad \alpha_{MPPM,MF}(\gamma) = e^{MPPM \times \Delta t} - 1.$$

3.2.3.5 Nonparametric Model

All previously described parametric measures are likely to suffer from the benchmark choice problem, since they do not correctly price passive portfolio returns, according to numerous asset pricing tests in the literature. Furthermore, their evaluation may change across models and other methodological choices (Lehmann and Modest, 1987, Grinblatt and Titman, 1994, Fama, 1998, and Ahn, Cao and Chrétien, 2009). To alleviate these problems, Chen and Knez (1996) develop the LOP performance measure, a nonparametric measure that prices passive portfolio returns by construction. Specifically, its SDF has the minimum volatility among SDFs that satisfy the law-of-one-price condition and corresponds to a linear function of the passive portfolio returns:

$$(22) \qquad m_{LOPt} = \mathbf{a}' \mathbf{R_{Kt}}.$$

The LOP SDF also represents the first part of the BC SDF introduced previously. Comparing both models can thus lead to a better understanding of the sources of disagreement that are associated with the second part of the BC SDF.

3.3 Methodology

In this section, we present the methodology for estimating and comparing the best clientele and candidate alphas.

3.3.1 Joint Estimation of the Performance Models

The main benefit of the methodology is that we perform a joint estimation of the alphas from the performance measures, leading to proper statistical inferences for the null and alternative hypotheses of section 3.2.1. The estimation is done by mapping the problems into moments and using the generalized method of the moments (GMM) of Hansen (1982) for estimation and testing. For a sample of T observations, we rely on the following basic set of moments:

$$
(23) \qquad \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}})\mathbf{R_{Kt}}] - \mathbf{1} = 0,
$$

$$
(24) \qquad \frac{1}{T}\sum_{t=1}^{T}[(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}})\mathbf{R_{Kt}}] = 0,
$$

$$
(25) \qquad \frac{1}{T}\sum_{t=1}^{T}[\bar{m}_{BCt}]^2 - \frac{\left(1+\bar{h}^2\right)}{R_F{}^2} = 0,
$$

$$
(26) \qquad \frac{1}{T}\sum_{t=1}^{T}[\bar{m}_{BCt}R_{MFt}] - 1 - \bar{\alpha}_{BC,MF} = 0,
$$

$$
(27) \qquad \frac{1}{T}\sum_{t=1}^{T}[m_{\varphi t}(\boldsymbol{\theta})R_{MFt}] - 1 - \alpha_{\varphi,MF}(\boldsymbol{\theta}) = 0,
$$

$$(28) \qquad \frac{1}{T}\sum_{t=1}^{T}\left[m_{\varphi t}(\boldsymbol{\theta})R_{Ft}\right] - 1 = 0.$$

Equations (23) to (26) represent the moments required to estimate the BC alpha. In total, it requires the estimation of $2K + 2$ parameters, where K is the number of passive portfolios. The K moments in equation (23) allow for the estimation of the LOP SDF, $m_{LOPt} = \mathbf{a}'\mathbf{R_{Kt}}$. It ensures that it correctly prices (on average) the K passive portfolio returns. The K moments in equation (24) represent orthogonality conditions between the replication error term, $w_t = R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}}$, and passive portfolio returns. These conditions are needed to estimate the coefficients $\mathbf{c}$ in the best replicating payoff $\mathbf{c}'\mathbf{R_{Kt}}$. The moment in equation (25) imposes the no-good-deal condition to estimate the parameter v, which is restricted to be positive to obtain an upper bound on performance. In this moment, $\overline{m}_{BCt} = m_{LOPt} + vw_t = \mathbf{a}'\mathbf{R_{Kt}} + v(R_{MFt} - \mathbf{c}'\mathbf{R_{Kt}})$ and $\overline{h}$ is an exogenously specified maximum Sharpe ratio. R_F represents a risk-free rate equivalent and is set to one plus the average one-month Treasury bill return in our sample, which is 0.3393%.[3] Finally, using the estimated BC performance SDF, $\overline{m}_{BCt}$, we obtain the BC alpha using the moment specified by equation (26).

The other two equations are useful for estimating the candidate alphas. Equation (27) is the moment for the estimation of $\alpha_{\varphi,MF}(\boldsymbol{\theta})$, the alpha associated with candidate SDF $m_{\varphi t}(\boldsymbol{\theta})$. Equation (28) is a moment imposed for all candidate parametric SDFs to force them to correctly price the risk free return. Dahlquist and Söderlind (1999) show the importance of fixing the mean of SDFs to an economically sound value in estimating SDF alphas. When necessary, this basic set of moments is augmented by the additional moments needed to estimate the parameters of candidate parametric SDFs.

For linear factor models, we add following moments:

---

[3] For consistency, we also include the one-month Treasury bill return as one of the K passive portfolio returns, so that the estimated mean SDF is similar to $1/R_F$.

$$(29) \qquad \frac{1}{T} \sum_{t=1}^{T} [m_{\varphi t}(\boldsymbol{\theta}) \mathbf{f_t}] - \mathbf{P_f} = 0.$$

These F moments allow for the estimation of the candidate linear factor SDF, $m_{\varphi t}(\boldsymbol{\theta}) = \omega_0 + \boldsymbol{\omega_1'} \mathbf{f_t}$, by ensuring that it correctly prices the F factors. $\mathbf{P_f}$ is a $F \times 1$ vector of factor prices, with $P_f = 1$ when the factor is a gross return and $P_f = 0$ when the factor is an excess return. Taken together, equations (28) and (29) lead to the estimation of any candidate linear factor with a just identified system. For example, the two parameters of the CAPM, $\boldsymbol{\theta} = \{a_0, a_{MKT}\}$, are estimated by correctly pricing the risk-free return and the market portfolio return. This strategy yields SDF alpha estimates for linear factor models that are closely related to Jensen-type alphas from regressions with the same factors. Similarly, for conditional linear factor models, we augment the basic set of moments with the following moments to account for the information variables $\mathbf{Z_{t-1}}$:

$$(30) \qquad \frac{1}{T} \sum_{t=1}^{T} [m_{C\varphi t}(\boldsymbol{\theta}) \mathbf{f_t}] - \mathbf{P_f} = 0,$$

$$(31) \qquad \left(\frac{1}{T} \sum_{t=1}^{T} [m_{C\varphi t}(\boldsymbol{\theta}) \mathbf{f_t}] - \mathbf{P_f} \right) \otimes \mathbf{Z_{t-1}} = 0.$$

For consumption-based models, instead of estimating the curvature parameter γ and the habit level parameter μ, we specify them directly by choosing relevant parameter values from the literature. Specifically, we follow Chrétien (2012, Table 3.6) and use $\gamma = \{2, 4, 6\}$ and $\mu = \{0.8, 0.9\}$. To again ensure an economically relevant mean SDF, we estimate the time discount factor β with equation (28) so that the resulting SDF correctly prices the risk free return.

For the manipulation-proof performance measure, we follow Goetzmann, Ingersoll, Spiegel and Welch (2007) and estimate alphas for risk aversion $\gamma = \{2, 3, 4\}$. As described

in section 3.2.3.4, the MPPM alpha does not come from a SDF model, so the moments in equations (27) and (28) are not necessary. Instead, we estimate it by exploiting the variability inherent in the time series average within the MPPM using the following moment:

(32)
$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{R_{MFt}}{R_{Ft}}\right]^{1-\gamma} - \left[1 + \alpha_{MPPM,MF}(\gamma)\right]^{1-\gamma} = 0.$$

To our knowledge, this paper is the first to propose an estimation strategy for MPPM alpha that allows for statistical inferences on the significance of the performance values.

Finally, for the nonparametric LOP performance measure, we use the SDF $m_{LOPt} = \mathbf{a}'\mathbf{R_{Kt}}$ estimated from the K moments in equation (23) to obtain the corresponding LOP alpha $\alpha_{LOP,MF}$. Hence, we do not need the moment in equation (28) in the system.

In all estimation cases, the systems are just identified because the number of parameters equals the number of moments. Hence, parameter estimates are not influenced by the choice of the weighting matrix in GMM. Although our procedure estimates alpha separately for each fund, Farnsworth, Ferson, Jackson and Todd (2002) demonstrate that estimating this system for one fund at a time produces the same point estimates and standard errors for alpha as a system that includes an arbitrary number of funds. The statistical significance of the parameters is assessed with the asymptotic properties of GMM derived by Hansen (1982).[4] The standard errors are adjusted for conditional heteroskedasticity and serial correlation using the method of Newey and West (1987) with two lags.[5]

---

[4] Chrétien and Kammoun (2015) investigate the finite sample properties of BC alpha estimates and find that inferences are generally robust to finite sample issues.

[5] Two lags account for the small but significant serial correlation in returns of some equity mutual funds (that might be invested in thinly traded stocks). Chrétien and Kammoun (2015) show that t-statistics for BC SDF alphas are similar when using no lag or four lags.

3.3.2 Passive Portfolios and Maximum Sharpe Ratio Choice

To implement the best clientele performance measure, two choices are particularly important: the passive portfolio returns $\mathbf{R_{Kt}}$ and the maximum Sharpe ratio $\bar{h}$. Chrétien and Kammoun (2015) analyze extensively the impact of these choices in the context of equity mutual fund performance evaluation. For the first choice, using sets of passive portfolio returns based on either ten industry portfolios, six style portfolios or the market portfolio, they find that BC alphas are relatively similar across the three sets. Although their main results rely on the ten industry portfolio set, they show that their conclusions are not sensitive to this choice. Based on this assessment, this paper selects the risk free rate and the returns on ten industry portfolios as passive portfolio returns. Details on these passive portfolios will be provided in the data section.

For the second choice, Chrétien and Kammoun (2015) present the literature on selecting an exogenous maximum Sharpe ratio, including Ross (1976), MacKinlay (1995) and Cochrane and Saá-Requejo (2000). They conclude that the ratio is somewhat subjectively specified, but that guidance from the literature leads to a selection that implies adding the market Sharpe ratio, $hMKT$, to the optimal Sharpe ratio attainable from the passive portfolios, h^*, or $\bar{h} = h^* + hMKT$. More conservatively, they also consider a maximum Sharpe ratio of $\bar{h} = h^* + 0.5hMKT$, which adds half of the market Sharpe ratio, as well as six other sensible choices. Based on the literature and their empirical analysis, they find that a value of $\bar{h} = h^* + 0.5hMKT$ appears the most relevant choice. In this paper, we follow this recommendation.

To implement it, we use $hMKT = 0.1262$, a value equal to the monthly market Sharpe ratio in our sample. The ten industry passive portfolio set yields a full-sample monthly optimal Sharpe ratio of $h^* = 0.258$, which gives a maximum Sharpe ratio of $\bar{h} = 0.321$ for funds with a full sample of observations. Adjusting h^* for the bias investigated by Ferson and Siegel (2003),[6] we obtain an adjusted h^* of 0.177 and an adjusted $\bar{h}$ of 0.240.

---

[6] Ferson and Siegel (2003) show that the sample optimal Sharpe ratio is biased upward when the number of basis assets (*K*) is large relative to number of observations (*T*). They propose a bias correction to obtain an

This value is similar to the maximum Sharpe ratios advocated by Ross (1976), MacKinlay (1995) and Cochrane and Saá-Requejo (2000), who respectively propose a monthly Sharpe ratio of 0.25, a squared annual Sharpe ratio of 0.6 (approximately equivalent to a monthly Sharpe ratio of 0.224) and an annual Sharpe ratio of 1 (approximately equivalent to a monthly value of 0.289).

3.3.3   Hypothesis Testing on Alphas

We provide numerous cross-sectional statistics to summarize the empirical results. First, we study the distribution of alpha estimates for each model. Specifically, we present the mean, standard deviation and selected percentiles of the distributions of estimated alphas and their corresponding t-statistics, computed as $t_{i,n} = \hat{\alpha}_{i,n}/\hat{\sigma}_{\hat{\alpha}_{i,n}}$, where $\hat{\alpha}_{i,n}$ is the estimated alpha for fund n and $\hat{\sigma}_{\hat{\alpha}_{i,n}}$ is its Newey-West standard error, for any performance measure $i = \{BC, \varphi\}$. We also present the following t-statistics to test for the hypothesis that the cross-sectional mean of estimated alphas is equal to zero:

$$(33) \qquad\qquad t_i = \frac{\hat{E}(\hat{\alpha}_i)}{\hat{\sigma}_{\hat{E}(\hat{\alpha}_i)}},$$

with:

$$(34) \qquad\qquad \hat{E}(\hat{\alpha}_i) = \frac{1}{N} \sum_{n=1}^{N} \hat{\alpha}_{i,n},$$

$$(35) \qquad\qquad \hat{\sigma}_{\hat{E}(\hat{\alpha}_i)} = \frac{\hat{\sigma}(\hat{\alpha}_{i,n})}{\sqrt{N}} \sqrt{1 + (N-1)\rho_{\hat{\alpha}_{i,n},\hat{\alpha}_{i,m}}},$$

---

adjusted optimal Sharpe ratio given by: $Adjusted\ h^* = \sqrt{\frac{(h^*)^2\ (T-K-2)}{T} - \frac{K}{T}}$. For our full sample, with the set of passive portfolios based on ten industry portfolios, $T = 336$ and $K = 11$.

where:

$$(36) \qquad \hat{\sigma}(\hat{\alpha}_i) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{\alpha}_{i,n}^2 - \hat{E}(\hat{\alpha}_i)^2)},$$

and N is the number of funds. This test assumes that the cross-sectional distribution of the N alphas obtained from any performance measure $i = \{BC, \varphi\}$ is multivariate normal with a mean of zero, a standard deviation $\hat{\sigma}(\hat{\alpha}_i)$ equal to the observed cross-sectional standard deviation of the alpha estimates, and a constant correlation between any two alpha estimates of $\rho_{\hat{\alpha}_{i,n},\hat{\alpha}_{i,m}}$, which we set at 0.044. This value is taken from Barras, Scaillet and Wermers (2010, p. 193) and Ferson and Chen (2015, Appendix, p. 62), who discuss the cross-sectional dependence in performance among funds, adjusted for data overlap, in their fund samples (which are comparable to ours).

Second, we study the disagreement between the BC alphas and the candidate alphas. We report the difference in means of estimated alphas, $\hat{E}(\hat{\bar{\alpha}}_{BC}) - \hat{E}(\hat{\alpha}_\varphi)$, the difference in standard deviations of estimated alphas, $\hat{\sigma}(\hat{\bar{\alpha}}_{BC}) - \hat{\sigma}(\hat{\alpha}_\varphi)$, as well as the correlation between estimated alphas, $\rho_{\hat{\bar{\alpha}}_{BC},\hat{\alpha}_\varphi}$, for each candidate model φ. We also present the following t-statistics to test for the hypotheses that these differences are respectively equal to zero:

$$(37) \qquad t_{\hat{E}(\hat{\bar{\alpha}}_{BC})-\hat{E}(\hat{\alpha}_\varphi)} = \frac{\hat{E}(\hat{\bar{\alpha}}_{BC}) - \hat{E}(\hat{\alpha}_\varphi)}{\sqrt{\hat{\sigma}_{\hat{E}(\hat{\bar{\alpha}}_{BC})}^2 + \hat{\sigma}_{\hat{E}(\hat{\alpha}_\varphi)}^2 - 2\rho_{\hat{\bar{\alpha}}_{BC},\hat{\alpha}_\varphi}\hat{\sigma}_{\hat{E}(\hat{\bar{\alpha}}_{BC})}\hat{\sigma}_{\hat{E}(\hat{\alpha}_\varphi)}}},$$

$$(38) \qquad t_{\hat{\sigma}^2(\hat{\bar{\alpha}}_{BC}) - \hat{\sigma}^2(\hat{\alpha}_\varphi)} = \frac{\left(\hat{\sigma}^2(\hat{\bar{\alpha}}_{BC}) - \hat{\sigma}^2(\hat{\alpha}_\varphi)\right)\sqrt{N}}{\sqrt{4\hat{\sigma}^2(\hat{\bar{\alpha}}_{BC})\hat{\sigma}^2(\hat{\alpha}_\varphi)\left(1 - \rho^2_{\hat{\bar{\alpha}}_{BC}, \hat{\alpha}_\varphi}\right)}}.$$

Details and literature on these tests can be found in Sheskin (1997, Test 12). We finally provide the proportions of the estimated candidate alphas that are smaller, significantly smaller, larger and significantly larger than the associated BC alphas. As discussed previously, the system of moments jointly estimate the BC and candidate alphas. This joint estimation leads to a direct test on the equality of alphas that properly accounts for the correlation between estimates.

3.4 Data and Summary Statistics

3.4.1 Mutual Fund Returns

The sample of funds includes actively-managed open-ended U.S. equity mutual funds from January 1984 to December 2012. The source is the CRSP Survivor-Bias-Free US Mutual Fund Database. We focus on U.S. equity funds by excluding bond and money market funds, balanced funds, international funds and funds not strongly invested in equity securities.[7] We exclude index funds and keep the funds only if they hold between 80% and 105% in common stocks.[8] Finally, we use the database variable "open to investors" to consider only open-ended mutual funds.

To mitigate survivorship bias and selection bias, we choose 1984 as the starting year, as suggested by Elton, Gruber and Blake (2001) and Fama and French (2010). To address back-fill and incubation biases, we follow Elton, Gruber and Blake (2001) and Kacperczyk, Sialm and Zheng (2008) and Evans (2010). We consider only observations

---

[7] Following Kacperczyk, Sialm and Zheng (2008), we identify open-ended U.S. equity mutual funds by policy code CS, Strategic Insight objective codes AGC, GMC, GRI, GRO, ING or SCG, Weisenberger objective codes G, G-I, AGG, GCI, GRO, LTG, MCG or SCG, and Lipper objective codes EIEI, EMN, LCCE, LCGE, LCVE, MATC, MATD, MATH, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE or SCVE.
[8] We identify index funds by finding the word "index" in their name and by Lipper objective codes SP and SPSP.

after the organization date and we require that funds have total net assets (TNA) superior to $15 million in the first year of entering the database. We also eliminate funds without a name and funds that omit to report their organization date. Finally, as Barras, Scaillet and Wermers (2010), we exclude funds that have fewer than 60 observations for estimation purposes.[9] Considering these sampling criteria, we obtain a final sample of 2786 actively-managed open-ended U.S. equity mutual funds.

3.4.2 Other Variables

As passive portfolios, we use the risk-free rate plus ten industry portfolios taken from Kenneth R. French's website. Industry classifications are consumer nondurables (NoDur), consumer durables (Durbl), manufacturing (Manuf), energy (Enrgy), high technology (HiTec), telecommunication (Telcm), shops (Shops), healthcare (Hlth), utilities (Utils) and other sectors (Others). The CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks (MKT) and the risk-free rate (RF) are taken from the CRSP database.

For the Fama-French and Carhart linear factor models, we take monthly returns on SMB, HML and MOM from Kenneth R. French's website. For the Ferson-Schadt model, we use returns on the S&P 500 index for large stocks (LS), on ninth and tenth NYSE market value deciles for small stocks (SS), on (approximatively 20-year) U.S. Treasury bonds for long-term government bonds (LTGB), on the Merrill Lynch High Yield Composite Index for low-grade corporate bonds (LGCB), and on one-month U.S. Treasury bills for the risk free asset (RF). The data are from CRSP except for LGCB, which comes from the Federal Reserve Economic Database (FRED).

For the conditional linear factor models, we consider the lagged values of four public information variables commonly used in the literature and first introduced by Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988) and Fama and French (1988, 1989). We use the dividend yield of the S&P 500 Index (DIV) from

---

[9] The 60-month screen ensures that we have enough observations to obtain reliable GMM estimates. The small survivorship bias it introduces should have little effect on the comparison between BC and candidate alphas. Furthermore, Barras, Scaillet and Wermers (2010) and Chrétien and Kammoun (2015) find that their performance findings are similar when using a 36-month screen instead of a 60-month screen.

Datastream, computed as the difference between the log of the twelve-month moving sum of dividends paid on the S&P 500 and the log of its lagged value; the yield on three-month U.S. Treasury bills (YLD) from FRED; the term spread (TERM), which is the difference between the long-term yield on U.S Treasury bonds (from Datastream) and the yield on three-month U.S. Treasury bills; and the default spread (DEF), which is the difference between BAA- and AAA-rated corporate bond yields from FRED.

For the consumption-based models, we proxy for aggregate per capita consumption using the seasonally-adjusted personal consumption expenditures on non-durables and services, adjusted by their respective consumption deflator, and divided by the resident population. For the price level, we take the non-seasonally-adjusted consumer price index. The data are from FRED.

3.4.3   Summary Statistics

Table 3.1 shows monthly summary statistics for the mutual fund returns (panel A), and for the factors, information variables and passive portfolio returns (panel B). In panel A, the mean mutual fund return is 0.734% with a standard deviation of 0.301%. The Sharpe ratios average 0.086 across funds, with values from -0.464 to 0.379. In panel B, the return factors have means from 0.068% (for SMB) to 1.028% (for SS), with standard deviations from 2.584% (LGCB) to 6.542% (for SS). Their Sharpe ratios vary from 0.022 (for SMB) to 0.322 (for LGCB). Consumption growth (CG) has a mean of 0.137% and a standard deviation of 0.285%. The information variables have means of 2.465% (for DIV), 4.126% (for YLD), 1.942% (for TERM) and 1.026% (for DEF). The industry portfolios have mean returns from 0.831% (for consumer durables) to 1.171% (for consumer nondurables), with standard deviations from 3.995% (for Utils) to 7.226% (for HiTec).

3.5   Empirical Results

This section presents the empirical results. We first examine the standard deviation of estimated SDFs in our sample. Then, we provide results for best clientele (BC) alphas and their closely related LOP alphas to establish the importance of investor disagreement within

the BC measure. Finally, we document the performance and disagreement associated with the various candidate parametric models under consideration.

3.5.1 Stochastic Discount Factors for Performance Evaluation

Our disagreement results are based on a comparison of performance evaluation obtained from BC SDFs with the one obtained from parametric SDFs. As discussed previously, all SDFs have similar means because they are required to correctly price the risk free return. A useful way to understand their differences is then to examine their volatility using the diagnostic tool developed by Hansen and Jagannathan (1991). Figure 3.1 provides a comparison of SDF standard deviations for unconditional linear factor models (figure 3.1a), conditional linear factor models (figure 3.1b), power utility models (figure 3.1c) and habit-formation preference models (figure 3.1d). The figure does not illustrate the MPPM because its alpha does not come from a SDF model.

In each figure, two lines indicate lower and upper volatility bounds for SDFs that are admissible under the law-of-one-price and no-good-deal conditions. The lower bound is the volatility restriction of Hansen and Jagannathan (1991), who show that admissible SDFs need a minimum volatility to correctly price passive portfolios. This minimum volatility SDF is also behind the LOP performance measure of Chen and Knez (1996), so the continuous line is given by $\sigma(m_{LOPt}) = h^*/R_F$. The upper bound is the maximum volatility restriction implied by the no-good-deal bounds of Cochrane and Saá-Requejo (2000), who demonstrate that ruling out good deals is equivalent to assuming a maximum SDF volatility. In our setup, the maximum volatility SDF is behind the BC performance measure, so the dashed line is equal to $\sigma(\bar{m}_{BCt}) = \bar{h}/R_F$.

All candidate SDFs with volatility outside the two lines are not admissible under the law-of-one-price and no-good-deal conditions, and thus could generate problematic alphas. Figure 3.1 reveals that candidate SDFs vary greatly with respect to their volatility. In figure 3.1a, the CAPM and FF SDFs do not have sufficient variability to be admissible, but the CARHART and FS SDFs are within the bounds. In figure 3.1b, the conditional versions of these SDFs have standard deviations that are too high, except for the CCAPM SDF. In figure 3.1c, as first illustrated by Hansen and Jagannathan (1991), POWER SDFs do not

have sufficient variability for reasonable values of the risk aversion parameter. Figure 3.1d shows that HABIT SDFs have higher volatility, almost reaching the lower bound when $\gamma = 6$ and $\mu = 0.9$, a finding in line with Cochrane and Hansen (1992). Overall, our candidate models provide a diversity of SDF volatilities, with values that are too low, well specified or too high, depending on the model. The next subsections examine whether this diversity leads to notable differences in their performance evaluation.

3.5.2 Best Clientele and LOP Alphas

Before turning to the results for different classes of parametric models, we start by presenting the alphas for the BC performance measure, which serve as a basis for comparison, and the alphas for the nonparametric LOP performance measure. Section 3.2.2 shows that the LOP SDF represents the first part of the BC SDF. Comparing both models is useful to better understand the sources of disagreement that are associated with the second part of the BC SDF, and hence establishes the importance of investor disagreement within the BC measure. As described in section 3.3.2, the measures use the returns on ten industry portfolios and the risk-free asset as passive portfolios. Also, the BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$, where h^* is the monthly optimal Sharpe ratio of the passive portfolios during the estimation period of a mutual fund, and $hMKT = 0.1262$ is the monthly Sharpe ratio of the CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks.

Table 3.2 presents the results by reporting the cross-sectional statistics and tests introduced in section 3.3.3. Specifically, this table as well as tables 3 to 7 include the following contents. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated monthly alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas. Panel B looks at the disagreement between BC alphas and candidate alphas (which are LOP alphas in table 3.2). On the left side of the panel, it reports differences in mean alphas (Mean Diff) and standard deviations of alphas (SD Diff), along with their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of

candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$), and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. All statistics are in percentage except the t-statistics.

Panel A of table 3.2 shows that the cross-sectional distribution of the estimated BC alphas has a mean of 0.236% (t-stat = 3.35) and standard deviation of 0.334%. The distribution of the BC t-statistics indicates that more than 75% of funds have positive alphas and approximately 25% of funds have significantly positive alphas. These findings are consistent with mutual funds providing added value to their best potential clienteles, as documented by Chrétien and Kammoun (2015). By comparison, the LOP alphas have a mean of -0.179% (t-stat = -3.14) and a standard deviation of 0.271%. More than 75% of funds have negative performance, significantly so for approximately 30% of funds. By eliminating the part of the BC SDF that accounts for the disagreement from the most favorable investors, the LOP SDF leads to a negative evaluation similar to the one commonly found in the mutual fund performance literature.

Panel B of table 3.2 confirms that the disagreement between both models is significant. The BC alphas are always greater than their corresponding LOP alphas *by construction*. However, their mean difference in alphas of 0.415% (t-stat =8.18) is economically large. Furthermore, this value is comparable to the magnitude of investor disagreement documented by Ferson and Lin (2014, Table III), who obtain bounds from 0.21% to 0.38% for the expected disagreement with traditional alphas for various benchmark returns. Even though the LOP SDF correctly prices passive portfolios and hence does not suffer from the benchmark choice problem, it suffers from the misrepresentation problem because it gives a severe or pessimistic fund evaluation compared with the one from the most favorable clienteles. The difference in standard deviations of alphas between the measures is also large (at 0.064%, t-stat =15.78), indicating that the BC SDF leads to a greater cross-sectional variability in alpha estimates. This finding is consistent with the BC measure allowing for higher disagreement on funds with returns difficult to span, leading to a non-constant disagreement across funds. The correlation of 0.70 between the BC and LOP alphas also indicates that both measures provide relatively different performance values due to the disagreement part of the BC SDF.

3.5.3 Alphas and Disagreement for the Unconditional Linear Factor Models

Table 3.3 presents the results for unconditional linear factor models, the most commonly used models in performance evaluation. In panel A, the alphas have means of -0.068% for the CAPM, -0.087% for the FF model, -0.120% for the CARHART model and -0.098% for the FS model. The standard deviations vary from 0.245% to 0.275%. Although a majority of funds have negative alphas, the mean alphas are only significantly different from zero for the FS model (at the 10% level, t-stat = -1.77) and the CARHART model (at the 1% level, t-stat = -2.32). The performance evaluation for the representative investors behind these models is not as negative as the one from the LOP measure presented previously.

As panel B shows, these representative investors nevertheless disagree significantly with the best clienteles. The mean differences in alphas are equal to 0.304% (t-stat = 5.50) for the CAPM, 0.323% (t-stat = 5.71) for the FF model, 0.356% (t-stat = 5.31) for the CARHART model, and 0.334% (t-stat = 6.17) for the FS model. These disagreement values are statistically and economically significant and once again comparable to the results of Ferson and Lin (2014). Similar to the LOP measure, linear factor models also significantly reduce the dispersion of fund alphas compared with the BC measure. They produce alphas that have smaller cross-sectional standard deviations and lower correlations with the BC alphas than the LOP alphas, particularly for the CARHART model. On a fund-by-fund basis, they generate alphas that are smaller than the BC alphas for more than 90% of funds and significantly smaller for approximately 50% of them. Although the SDF volatility bounds in figure 3.1a differentiate the CAPM and FF SDFs (not admissible) from the CARHART and FS SDFs (admissible), the alphas for all four models show little differences. They are almost always below the BC alpha admissibility bounds and oftentimes suffer from the misrepresentation problem.

3.5.4 Alphas and Disagreement for the Conditional Linear Factor Models

Table 3.4 presents results for conditional linear factor models, which have been used extensively in performance evaluation since their introduction by Ferson and Schadt (1996). The findings are qualitatively similar to the ones for their unconditional counterparts. Their average alphas are negative, significantly so at the 5% level only for the

112

CCARHART model (t-stat = -2.60). Their mean differences in alphas vary from 0.288 (for CFS, t-stat = 5.42) to 0.374 (for CCARHART, t-stat = 5.28), indicating a large and statistically significant disagreement. They also generate a significant decrease in the cross-sectional dispersion of fund alphas compared with the BC measure. Although the volatility bounds in figure 3.1b show that conditional models have SDF standard deviations that are too high, the alphas for all four models are almost always below the BC alpha bounds, and significantly so for approximately 50% of funds. Hence, such as their unconditional versions, these models suffer from the misrepresentation problem because they give alphas that likely undervalue the funds for their most favorable clienteles.

Overall, the results for unconditional and conditional linear factor models show that there is an economically and statistically significant disagreement between the performance evaluation for representative investors and the one for best clienteles. This disagreement leads the models to a somewhat severe fund evaluation compared with the one for the most favorable clienteles.

3.5.5 Alphas and Disagreement for the Consumption-Based Models

Tables 3.5 and 3.6 present the results for consumption-based models. The performance results across various specifications of the curvature parameter γ and habit level parameter μ are remarkably similar, with almost the same correlation (estimates ≈ 0.63) between their alphas and the BC alphas. In panel A of table 3.5, the cross-sectional distributions of alphas for POWER models have means and standard deviations of 0.429% (t-stat = 6.89) and 0.230% when $\gamma = 2$, 0.423% (t-stat = 6.80) and 0.296% when $\gamma = 4$, and 0.417% (t-stat = 6.69) and 0.296% when $\gamma = 6$. Based on the power utility models, more than 90% of funds have positive alphas and approximately 25% of funds have significantly positive alphas. In panel A of table 3.6, the results are more positive for the habit-formation preference models. For example, the mean alphas range from 0.448% (t-stat = 7.25), when $\gamma = 2$ and $\mu = 0.8$, to 0.542% (t-stat = 8.02), when $\gamma = 6$ and $\mu = 0.9$.

Panel B of tables 3.5 and 3.6 confirms that the consumption-based alphas are generally not admissible because their mean values are significantly higher than the upper bounds given by the BC alphas. The mean differences in alphas are less than -0.180% for

all specifications. Also, on a fund-by-fund basis, the consumption-based alphas are larger than the BC alphas for more than 80% of funds. In addition to having SDF volatilities that are too low, as illustrated by figures 1c and 1d, the consumption-based models rely on a proxy of per capita consumption growth that has a low correlation with equity returns, as discussed in the equity premium puzzle literature. The joint effects of low volatility and low correlation lead consumption-based SDFs to insufficiently discount the mutual fund gross returns, resulting in alphas that are too high to be admissible.

3.5.6   Alphas and Disagreement for the Manipulation Proof Performance Measure

Table 3.7 gives the results for the MPPM with risk aversion parameter $\gamma = \{2, 3, 4\}$, following Goetzmann, Ingersoll, Spiegel and Welch (2007). For this measure, it is useful to describe the performance results in more details because this paper is the first to estimate monthly effective MPPM SDF alphas, their standard errors and their t-statistics for a large sample of equity mutual funds. To our knowledge, the MPPM literature has focused mostly on hedge funds and has not provided an estimation strategy that allows for statistical inferences on the significance of the performance values. The only published exception for mutual funds is Ferson and Lin (2014), who presents MPPM alphas with $\gamma = 2.649$, but do not provide standard errors or t-statistics.[10]

Panel A provides three interesting findings. First, the MPPM alphas are relatively sensitive to the choice of γ. They have means of 0.120% (t-stat = 1.40) when $\gamma = 2$, -0.050% (t-stat = -0.36) with $\gamma = 3$, and -0.228% (t-stat = -1.05) when $\gamma = 4$. These results can be interpreted in light of the role of γ in the MPPM. Intuitively, MPPM representative investors have increasing risk aversion and "aversion to manipulation" as γ increases. Our results show that when they have low concerns about risk and manipulation (i.e., $\gamma = 2$), their alphas are closer to the BC alphas. When they have medium concerns about risk and manipulation (i.e., $\gamma = 3$), their alphas are similar to the linear factor model alphas. When they have high concerns about risk and manipulation (i.e., $\gamma = 4$), their alphas are the most

---

[10] According to Ferson and Lin (2014, footnote 22), they appear to compute the continuously compounded version of the MPPM given by equation (20).

negative of all models, consistent with alphas being inflated by manipulation for investors who care the most about it.

Second, the MPPM alphas are difficult to estimate with precision. Looking at the distribution of the t-statistics on the right side of panel A, the proportions of funds with significant alphas at the 5% level are much lower for the MPPM than for other models. When $\gamma = 2$, only approximately 2% of funds have significantly positive alphas and fewer than 1% of funds have significantly negative alphas. When $\gamma = 3$, fewer than 1% of funds have significantly positive alphas and significantly negative alphas. When $\gamma = 4$, fewer than 1% of funds have significantly positive alphas and approximately 2.5% of funds have significantly negative alphas. Consistent with this imprecision, the mean alphas are not significantly different from zero for the three values of γ considered in this paper.

Third, the MPPM results in much larger cross-sectional standard deviations of fund alphas compared with the other models. (Panel B confirms that it produces a significantly larger dispersion of alphas than the BC measure.) In previous subsections, we found that the highest dispersion belongs to the BC SDF and attributed the finding to the large disagreement it produces for funds with returns difficult to span by the most favorable investors. The even larger MPPM alpha standard deviation raises the question of whether manipulation generates even more disagreement or whether it is the result of noisier MPPM alpha estimates. Ferson and Lin (2014) indirectly investigate the issue of disagreement with the MPPM and argue that manipulation is not likely to be the main source. The MPPM alpha distributions in panel A of table 3.7 provide evidence in support of their analysis. Looking at the interquartile range of alphas as an alternative measure, the MPPM alphas are less disperse than the BC alphas. In fact, for funds with alphas within the 5th and 95th percentiles, the MPPM generates dispersion in line with traditional parametric models. But the MPPM alpha distributions present a larger negative skewness, which is the reason for their larger standard deviations. For a small number of funds, manipulation-proofing their performance gives significantly lower alphas than their traditional or BC alphas. For these funds, it remains an open question whether the MPPM uncovers truly strong manipulation or whether their MPPM alpha estimates are simply very noisy.

Panel B of table 3.7 confirms that the disagreement between the MPPM and the BC measure is not statistically significant when $\gamma = 2$, but becomes important when $\gamma = 3$ and $\gamma = 4$. When $\gamma = 2$, the mean difference in alphas of 0.116% (t-stat = 1.24) is small and there is little evidence that the MPPM alphas are significantly different from the BC alphas. The MPPM with $\gamma = 2$, a specification in which the representative investors have relatively low concerns about risk and manipulation, thus gives alphas that are admissible and adequately represent the mean alpha for the most favorable clienteles. In the specifications where the representative investors are more concerned about risk and manipulation, the mean differences in alphas are equal to 0.286% (t-stat = 2.03) when $\gamma = 3$ and 0.464% (t-stat = 2.12) when $\gamma = 4$. These statistically significant disagreement values are once again comparable to the results of Ferson and Lin (2014). Furthermore, for all the MPPM specifications, the correlations between their alphas and the BC alphas are the lowest among all candidate models, with estimates ranging from 0.15 to 0.30. These low values can be mostly attributed to the high MPPM alpha standard deviation rather than the low covariance between the alphas. Finally, on a fund-by-fund basis, 69.8% and 84.0% of the MPPM alphas are smaller than the BC alphas when, respectively, $\gamma = 3$ and $\gamma = 4$. Although the noise in MPPM alpha estimates makes the proportions of significantly smaller alphas being less than 10%, our evidence suggests that the MPPM with $\gamma = 3$ or $\gamma = 4$ produces alphas that oftentimes suffer from the misrepresentation problem.

3.5.7 Assessing the Severity of the Misrepresentation Problem with Worst Clientele Alphas

Our analysis has thus far relied on a comparison between the best clientele and candidate alphas. Except for the MPPM measure with a low risk aversion parameter, we find that all candidate models are problematic. The unconditional and conditional linear factor models, the MPPM with a high risk aversion parameter and the LOP measure suffer from the misrepresentation problem. The consumption-based models suffer from the inadmissibility problem. The latter problem is a more damaging diagnosis for a performance measure because it indicates that the measure fails to meet the requirements for admissibility (most importantly, to correctly price passive portfolios) in such a significant way as to generate unreliable performance values. The misrepresentation problem is a less fundamental

problem that indicates when a measure provides pessimistic performance values in disagreement with the evaluation of the funds for their best clienteles.

However, it is possible that a measure suffers so badly from the misrepresentation problem as to render it inadmissible: an inadmissibility problem can also be diagnosed when a candidate alpha is smaller than the lower admissible performance bound, called the "worst clientele alpha" by Chrétien and Kammoun (2015). Given the presence of the misrepresentation problem for many candidate models we investigate, this section assesses the severity of the problem by comparing the candidate alphas with the worst clientele alphas.

From the solution to the lower bound problem derived by Cochrane and Saá-Requejo (2000), the worst clientele (WC) alpha is given by $\underline{\alpha}_{WC,MF} = E[\underline{m}_{WCt} R_{MFt}] - 1$, where the WC SDF is $\underline{m}_{WCt} = m_{LOPt} - vw_t$. Using the estimation strategy presented previously, with the same passive portfolios and allowing for the same maximum Sharpe ratio (i.e., $\bar{h} = h^* + 0.5hMKT$), we estimate the WC alphas and compute the disagreement between the WC and candidate alphas for five candidate models: the FF and CARHART linear factor models, their conditional versions (i.e., CFF and CCARHART) and the MPPM with risk aversion parameter $\gamma = 4$. Empirically, these parametric models generate the most negative mean alphas and hence suffer the most from the misrepresentation problem.

Table 3.8 presents the results, using a format similar to tables 3.2 to 3.7, with the performance results for the candidate models repeated from previous tables for ease of comparison. In panel A, the cross-sectional distribution of the WC alphas has a mean of -0.594% (t-stat = -7.26) and a standard deviation of 0.388%. As documented by Chrétien and Kammoun (2015), the worst clienteles value funds more negatively than previous evidence shows. Panel B looks at the disagreement between WC alphas and candidate alphas, and shows that the five candidate models do not suffer from an inadmissibility problem with respect to the lower bound. Their mean differences in alphas vary from -0.366% (t-stat = -2.11) for the MPPM with $\gamma = 4$ to -0.507 (t-stat = 9.27) for the FF model, indicating that the candidate alphas are significantly higher than the WC alphas. Nevertheless, except for the CFF model, the correlations between the candidate alphas and

the WC alphas are higher than their corresponding values computed with the BC alphas. This result suggests that the representative investors behind the candidate models are more related to the worst clienteles than to the best clienteles. On a fund-by-fund basis, the candidate alphas are larger than the WC alphas for more than 94% of funds, significantly so for at least 78% of funds for the linear factor models and 11.5% of funds for the MPPM. Overall, these findings confirm the admissibility of the candidate models that suffer from the misrepresentation problem, because their alphas are within the lower and upper admissible performance bounds.

3.6 Conclusion

This paper examines standard performance measures by comparing their alpha with the alpha from a performance measure that evaluates mutual funds from the point of view of their most favorable investors. This yardstick for comparison is termed "best clientele alpha" by Chrétien and Kammoun (2015), and we obtain it by estimating the upper admissible bound under the law-of-one-price and no-good-deal conditions developed by Cochrane and Saá-Requejo (2000).

Two alternative hypotheses are insightful in our setup. On one hand, an inadmissibility problem occurs when a candidate alpha is greater than the upper admissible bound that is the best clientele alpha. On the other hand, a misrepresentation problem occurs when a candidate alpha is lower than the best clientele alpha, because the candidate alpha provides a "severe" or "pessimistic" evaluation of the fund that does not adequately reflect the more useful evaluation from its best potential clienteles. The misrepresentation problem is also indicative of large investor disagreement in performance evaluation.

We conduct our investigation on twelve candidate models: four unconditional linear factor models (the CAPM and the Fama and French (1993), Carhart (1997) and Ferson and Schadt (1996) models), four conditional linear factor models (conditional versions of the previous four models), two consumption-based models (a power utility model and an external habit-formation model), the manipulation proof performance measure (MPPM) of Goetzmann, Ingersoll, Spiegel and Welch (2007) and the nonparametric LOP measure of Chen and Knez (1996). As we vary parameter choices for some models, we consider a total

of 21 specifications for the twelve models. We use a sample of 2786 actively-managed open-ended U.S. equity mutual funds with returns from 1984 to 2012 to perform our diagnosis of these models.

Among the empirical implementations of our candidate models, the case of the MPPM is particularly noteworthy. This paper is the first to estimate monthly effective MPPM SDF alphas, their standard errors and their t-statistics for a large sample of equity mutual funds. To our knowledge, no estimation strategy allowing for statistical inferences on the significance of the performance values exists in the MPPM literature. Our empirical results for the MPPM document three interesting findings. First, the MPPM alphas are relatively sensitive to the choice of risk aversion parameter so that when the MPPM representative investors have low concerns about risk and manipulation, their alphas tend to be positive, but when they have high concerns, their alphas are the most negative of all models, consistent with traditional alphas being inflated by manipulation from the point of view of investors who care the most about it. Second, the MPPM alphas are difficult to estimate with precision and the proportions of funds with significant alphas are much lower. Third, the MPPM generates a much larger cross-sectional standard deviation of fund alphas compared with the other models because its alpha distribution presents a large negative skewness. For a small number of funds, manipulation-proofing their performance gives significantly lower alphas than their traditional or best clientele alphas.

Finally, our comparison between candidate alphas and best clientele alphas shows that most models generally misrepresent the value of mutual funds for the most favorable clienteles. Specifically, the unconditional linear factor models, their conditional versions, the MPPM with high risk aversion parameter and the LOP measure give a severe but admissible evaluation of fund performance. The average performance disagreement between the best clienteles and the representative investors from these candidate models, measured by the mean difference between their alphas, vary from 0.283% to 0.464%. These economically and statistically significant disagreement values are comparable to the results of Ferson and Lin (2014). In opposite, alphas from consumption-based models are oftentimes not admissible because they are too high. Among all models, we find that a

MPPM with a low risk aversion parameter is the most appropriate in providing admissible alphas that reflect the value of funds for the best clienteles.

There are many avenues for future research relevant to our analysis. For example, the diagnostic tool proposed in this paper is sufficiently general that it could be applied to document the misrepresentation and inadmissibility problems of numerous other performance models, potentially using other types of investment portfolios. Future research could also identify and study the determinants of our performance disagreement measure to characterize funds for which investors strongly disagree. As discussed by Ferson (2010) and Ferson and Lin (2014), better understanding investor disagreement and developing clientele-specific performance measures are important challenges for future research. This paper makes sizeable contributions in that regard.

3.7   References

Abel, A.B., (1990), "Asset prices under habit formation and catching up with the Joneses", *American Economic Review*, 80, 38-42.

Ahn, D., Cao, H. H., Chrétien, S., (2009), "Portfolio performance measurement: a no arbitrage bounds approach"*, European Financial Management*, 15, 2, 298-339.

Bailey, W., Kumar, A., Ng, D., (2011), "Behavioral biases and mutual fund investors", *Journal of Financial Economics*, 102, 1-27.

Barber, B. M., Odean, T., Zheng, L., (2005), "Out of sight, out of mind: the effects of expenses on mutual fund flows", *Journal of Business*, 78, 2095-2120.

Barras, L., Scaillet, O., Wermers, R., (2010), "False discoveries in mutual fund performance : measuring luck in estimated alphas", *Journal of Finance*, 65, 1, 179-216.

Bergstresser, D., Chalmers, J., Tufano, P., (2009), "Assessing the costs and benefits of brokers in the mutual fund industry", *Review of Financial Studies*, 22, 4129-4156.

Campbell, J. Y., (1987), "Stock returns and the term structure", *Journal of Financial Economics*, 18, 373-399.

Campbell, J. Y., Shiller, R. J., (1988), "The dividend price ratio and expectations of future dividends and discount factors", *Review of Financial Studies*, 1, 3, 195-228.

Carhart, M. M., (1997), "On persistence in mutual fund performance", *Journal of Finance*, 52, 57-82.

Chen, Z., Knez, P. J., (1996), "Portfolio performance measurement: theory and applications", *The Review of Financial Studies*, 9, 2, 511-555.

Chrétien, S., (2012), "Bounds on the autocorrelation of admissible stochastic discount factors", *Journal of Banking and Finance*, 36, 1943-1962.

Chrétien, S., Coggins, F., d'Amours, F., (2015), "The performance of market timing measures in a simulated environment", Unpublished Working Paper, Laval University.

Chrétien, S., Kammoun, M., (2015), "Mutual fund performance evaluation and best clienteles", Unpublished Working Paper, Laval University.

Christoffersen, S., Evans, R. Musto, D., (2013), "What do consumers' fund flows maximize? Evidence from their brokers' incentives", *Journal of Finance*, 68, , 201-235.

Cochrane, J. H., Hansen, L. P., (1992), "Asset pricing explorations for macroeconomics", In: Blanchard, O., Fischer, S. (Eds.), *1992 NBER Macroeconomics Annual*, MIT Press, Cambridge.

Cochrane, J. H., Saá-Requejo, J., (2000), "Beyond arbitrage: good-deal asset price bounds in incomplete markets", *Journal of Political Economy,* 108, 79–119.

Cremers, M., Petajisto, A., Zitzewitz, E., (2013), "Should benchmark indices have alpha? Revisiting performance evaluation", *Critical Finance Review*, 2, 1-48.

Del Guercio, D., Reuter, J., (2014), "Mutual fund performance and the incentive to generate alpha", *Journal of Finance,* 69, 1673-1704.

Elton, E. J., Gruber, M. J., Blake, C. R., (2001), "A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases", *Journal of Finance*, 56, 2415–2430.

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

Fama, E. F., (1998), "Market efficiency, long-term returns and behavioral finance", *Journal of Financial Economics*, 49, 283-306.

Fama, E. F., French, K. R., (1988), "Permanent and temporary components of stock prices", *Journal of Political Economy*, 96, 2, 246-273.

Fama, E. F., French, K. R., (1989), "Business conditions and expected returns on stocks and bonds", *Journal of Financial Economics*, 25, 23-49.

Fama, E. F., French, K. R., (1993), "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, 33, 3-56.

Fama, E. F., French, K. R., (2010), "Luck versus skill in the cross-section of mutual fund returns", *Journal of Finance*, 65, 5, 1915-1948.

Farnsworth, H., Ferson, W., Jackson, D., Todd, S., (2002), "Performance evaluation with stochastic discount factors", *Journal of Business*, 75, 3, 473-503.

Ferson, W. E., (2010), "Investment performance evaluation", *Annual Review of Financial Economics*, 2, 207-234.

Ferson, W. E., Chen, J., (2015), "How many good and bad fund managers are there, really?", *Working Paper*, University of Southern California.

Ferson, W. E., Lin, J., (2014), "Alpha and performance measurement: the effect of investor disagreement and heterogeneity", *Journal of Finance*, 69, 4, 1565-1596.

Ferson, W. E., Schadt, R., (1996), "Measuring fund strategy and performance in changing economic conditions", *Journal of Finance*, 51, 425-462.

Ferson, W. E., Siegel, A. F., (2003), "Stochastic discount factor bounds with conditioning information", *Review of Financial Studies*, 16, 567-595.

Glosten, L., Jagannathan, R., (1994), "A contingent claims approach to performance evaluation", *Journal of Empirical Finance*, 1, 133-66.

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I., (2007), "Portfolio performance manipulation and manipulation-proof performance measures", *Review of Financial Studies*, 20, 5, 1503-1546.

Grinblatt, M., Titman, S., (1994), "A study of monthly mutual fund returns and performance evaluation techniques", *Journal of Financial and Quantitative Analysis*, 29, 419–444.

Hansen, L. P., (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029–1054.

Hansen, L. P., Jagannathan, R., (1991), "Implication of security market data for models of dynamic economies", *Journal of Political Economy*, 99, 2, 225-262.

Keim, D. B., Stambaugh, R. F., (1986), "Predicting returns in the stock and bond markets", *Journal of Financial Economics*, 17, 2, 357-390.

Kosowski, R., Timmermann, A., Wermers, R., White, H.L., (2006), "Can mutual fund "stars" really pick stocks? Evidence from a bootstrap analysis", *Journal of Finance,* 61, 2251-2295.

Kothari, S.P. and Warner, J. B., (2001), "Evaluating mutual fund performance", *Journal of Finance*, 56, 1985-2010.

Kacperczyk, M., Sialm, C., Zheng, L., (2008), "Unobserved actions of mutual funds", *Review of Financial Studies*, 21, 6, 2379-2416.

Lehmann, B. N., Modest, D. M., (1987), "Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons", *Journal of Finance,* 42, 233-66.

Malloy, C., Zhu, N., (2004), "Mutual fund choices and investor demographics", Unpublished Working Paper, University of California at Davis.

Nanda, V., Narayanan, M. P., Warther, V., (2000), "Liquidity, investment ability, and mutual fund structure", *Journal of Financial Economics*, 57, 417-443.

Nanda, V., Wang, Z. J., Zheng, L., (2009), "The ABCs of mutual funds: on the introduction of multiple class shares ", *Journal of Financial Intermediation*, 18, 329-361.

Newey, W. K., West, K. D., (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703–708.

Ross, S. A., (1976), "The arbitrage theory of capital asset pricing", *Journal of Economics Theory*, 13, 341-360.

Sheskin, D. J., (1997), Handbook of Parametric and Nonparametric Statistical Procedures, CRC Press.

Table 3.1: Summary Statistics

Table 3.1 presents summary statistics for the monthly data from January 1984 to December 2012. Panel A shows cross-sectional summary statistics (average (Mean), standard deviation (Std Dev) and selected percentiles) on the distributions of the average (Mean), standard deviation (SD), minimum (Min), maximum (Max) and Sharpe ratio (h) for the returns on 2786 actively-managed open-ended U.S. equity mutual funds. Panel B gives the average (Mean), standard deviation (SD), minimum (Min), maximum (Max) and Sharpe ratio (h) for the factors, information variables and passive portfolio returns. Factors include the Fama-French market, size and value factors (MKT-RF, SMB and HML), the Carhart momentum factor (MOM), the Ferson-Schadt four factors (large stock (LS), small stock (SS), long-term government bond (LTGB) and low-grade corporate bond (LGCB)) and consumption growth (CG). The information variables are lagged values of the dividend yield on the S&P500 Index (DIV), the yield on the three-month Treasury bills (YLD), the term spread (TERM) and the default spread (DEF). The passive portfolios include ten industry portfolios (consumer nondurables (NoDur), consumer durables (Dur), manufacturing (Manuf), energy (Enrgy), high technology (HiTec), telecommunication (Telcm), shops (Shops), healthcare (Hlth), utilities (Utils), and other industries (Other)), and the risk free asset (RF) based on the one-month Treasury bills. All statistics are in percentage except for the Sharpe ratios.

	Panel A: Mutual Fund Returns and Carhart Alphas						
	Mutual Fund Returns					Carhart Alphas	
	Mean	StdDev	Min	Max	h	α_{MF}	t-statistics
Mean	0.7338	5.3400	-19.9976	16.4080	0.0857	-0.1200	-0.7398
StdDev	0.3008	1.5632	5.6560	7.8772	0.0532	0.2456	1.2534
(t-stat)						(-2.321)	
99%	1.3677	10.3594	-5.5453	41.5517	0.1990	0.4231	2.2289
95%	1.1405	8.2353	-12.8253	32.5704	0.1593	0.2120	1.3169
90%	1.0453	7.2002	-14.4430	27.0565	0.1427	0.1271	0.8321
75%	0.9044	6.0719	-16.5652	18.5645	0.1178	0.0109	0.0731
Median	0.7464	5.0272	-19.4020	14.1103	0.0900	-0.1016	-0.7466
25%	0.5946	4.3865	-22.9302	11.4669	0.0613	-0.2214	-1.5149
10%	0.4232	3.8968	-26.3190	9.9918	0.0229	-0.3811	-2.2580
5%	0.2943	3.4811	-29.0886	9.0829	-0.0033	-0.5037	-2.7996
1%	-0.1139	1.6197	-36.9313	5.3715	-0.0827	-0.8329	-4.1538

Table 3.1: Summary Statistics (continued)

Panel B: Passive Portfolio Returns and Information Variables					
	Mean	StdDev	Min	Max	*h*
Industry Portfolios					
NoDur	1.1713	4.3493	-21.0300	14.7400	0.1962
Durbl	0.8311	7.0347	-32.8900	42.9200	0.0698
Manuf	1.0667	5.1203	-27.3200	17.7800	0.1420
Enrgy	1.1223	5.3691	-18.3900	19.1300	0.1459
HiTec	0.9338	7.2260	-26.1500	20.4600	0.0822
Telcm	0.9689	5.2612	-15.5600	22.1200	0.1199
Shops	1.0394	5.0898	-28.3100	13.3800	0.1375
Hlth	1.1140	4.7552	-20.4700	16.5400	0.1636
Utils	0.9444	3.9952	-12.6500	11.7600	0.1521
Other	0.8829	5.3165	-23.6800	16.1100	0.1024
Style Portfolios, Market Portfolio and Risk-Free Asset					
B/L	0.9403	4.7004	-23.1900	14.4500	0.1281
B/M	0.9705	4.5832	-20.3200	14.8500	0.1378
B/H	0.9279	5.2367	-24.4700	22.1600	0.1126
S/L	0.8037	6.7657	-32.3400	27.0200	0.0685
S/M	1.1221	5.2548	-27.5700	18.8700	0.1487
S/H	1.2282	6.2180	-28.0500	38.3900	0.1426
MKT	0.9174	4.5814	-22.5363	12.8496	0.1262
RF	0.3393	0.2166	0.0000	1.0000	-
Information Variables					
DIV	2.4649	0.9204	1.0800	4.9900	-
YLD	4.1264	2.6038	0.0100	10.4700	-
TERM	1.9419	1.1392	-0.5300	3.7600	-
DEF	1.0255	0.4046	0.5500	3.3800	-

Table 3.2: Performance Disagreement for the LOP measure

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the best clientele (BC) performance measure and the LOP performance measure of Chen and Knez (1996). The BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between BC alphas and LOP alphas (the candidate alphas). On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$) and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlation.

Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds				
	Performance		t-statistics	
	BC	LOP	BC	LOP
Mean	0.2360	-0.1789	1.0220	-1.2291
Std Dev	0.3344	0.2707	1.4333	1.4743
(t-stat)	(3.351)	(-3.138)		
99%	1.2708	0.3580	4.4955	2.0211
95%	0.7851	0.1787	3.3544	1.0473
90%	0.644	0.0967	2.7535	0.5435
75%	0.4225	-0.0364	1.9417	-0.2258
Median	0.1841	-0.1630	1.0652	-1.1398
25%	0.0223	-0.2854	0.1720	-2.1715
10%	-0.1029	-0.4501	-0.8232	-3.1627
5%	-0.1909	-0.5978	-1.4254	-3.6814
1%	-0.4353	-0.9247	-2.6052	-4.8763

Panel B: Performance Evaluation Disagreement			
Mean Diff	0.4149	% $\alpha_\varphi < \bar{\alpha}_{BC}$	100.00
(t-stat)	(8.178)	% $\alpha_\varphi sig < \bar{\alpha}_{BC}$	99.57
SD Diff	0.0637	% $\alpha_\varphi > \bar{\alpha}_{BC}$	0.00
(t-stat)	(15.777)	% $\alpha_\varphi sig > \bar{\alpha}_{BC}$	0.00
Corr	0.7019		

Table 3.3: Performance Disagreement for Unconditional Linear Factor Models

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the best clientele (BC) performance measure and the CAPM, FF, CARHART and FS unconditional linear factor models. The BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between BC alphas and linear factor model alphas (the candidate alphas). On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$) and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlations.

	Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds									
	Performance					t-statistics				
	BC	CAPM	FF	CARHART	FS	BC	CAPM	FF	CARHART	FS
Mean	0.2360	-0.0682	-0.0866	-0.1200	-0.0984	1.0220	-0.3805	-0.5515	-0.7398	-0.5960
Std Dev	0.3344	0.2747	0.2576	0.2456	0.2642	1.4333	1.1968	1.3061	1.2534	1.2820
(t-stat)	(3.351)	(-1.180)	(-1.597)	(-2.321)	(-1.769)					
99%	1.2708	0.4665	0.4354	0.4231	0.4132	4.4955	2.2944	2.3957	2.2289	2.0461
95%	0.7851	0.2957	0.2376	0.2120	0.2471	3.3544	1.4232	1.4457	1.3169	1.1640
90%	0.644	0.2170	0.1643	0.1271	0.1615	2.7535	1.0394	1.0263	0.8321	0.7944
75%	0.4225	0.0768	0.0431	0.0109	0.0386	1.9417	0.4052	0.3046	0.0731	0.2010
Median	0.1841	-0.0515	-0.0654	-0.1016	-0.0789	1.0652	-0.3109	-0.5098	-0.7466	-0.4792
25%	0.0223	-0.1843	-0.1847	-0.2214	-0.2069	0.1720	-1.1114	-1.3427	-1.5149	-1.2106
10%	-0.1029	-0.3454	-0.3376	-0.3811	-0.3588	-0.8232	-1.8900	-2.2267	-2.2580	-2.1713
5%	-0.1909	-0.4818	-0.4622	-0.5037	-0.4986	-1.4254	-2.4347	-2.6916	-2.7996	-2.8169
1%	-0.4353	-0.8637	-0.8416	-0.8329	-0.8184	-2.6052	-3.7564	-4.1733	-4.1538	-3.9760

	Panel B: Performance Evaluation Disagreement								
Mean Diff		0.3042	0.3226	0.3560	0.3344	% $\alpha_\varphi < \bar{\alpha}_{BC}$	97.81	97.56	93.36	99.28
(t-stat)		(5.503)	(5.706)	(5.309)	(6.170)	% $\alpha_\varphi sig < \bar{\alpha}_{BC}$	50.04	48.64	50.83	50.75
SD Diff		0.0597	0.0768	0.0888	0.0702	% $\alpha_\varphi > \bar{\alpha}_{BC}$	2.19	2.44	6.64	0.72
(t-stat)		(13.660)	(17.681)	(18.338)	(16.575)	% $\alpha_\varphi sig > \bar{\alpha}_{BC}$	0.00	0.00	0.04	0.00
Corr		0.6443	0.6159	0.4307	0.6530					

Table 3.4: Performance Disagreement for Conditional Linear Factor Models

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the best clientele (BC) performance measure and the CCAPM, CFF, CCARHART and CFS conditional linear factor models. The BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. The information variables used for conditional models are lagged values of the dividend yield on the S&P500 Index, the yield on three-month Treasury bills, the term spread and the default spread. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between BC alphas and conditional linear factor model alphas (the candidate alphas). On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$) and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlations.

| | **Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds** | | | | | | | | | |
| | **Performance** | | | | | **t-statistics** | | | | |
	BC	CCAPM	CFF	CCARHART	CFS	BC	CCAPM	CFF	CCARHART	CFS
Mean	0.2360	-0.0714	-0.0893	-0.1382	-0.0523	1.0220	-0.5031	-0.6974	-0.9863	-0.5257
Std Dev	0.3344	0.2741	0.2487	0.2521	0.3028	1.4333	1.2019	1.4016	1.4104	1.3307
(t-stat)	(3.351)	(-1.237)	(-1.705)	(-2.604)	(-0.820)					
99%	1.2708	0.5374	0.4640	0.4677	0.6996	4.4955	2.0894	2.3854	2.1929	2.2009
95%	0.7851	0.3008	0.2423	0.1955	0.3807	3.3544	1.3342	1.4461	1.3115	1.3723
90%	0.644	0.2066	0.1681	0.1166	0.2542	2.7535	0.9603	1.0373	0.8311	1.0337
75%	0.4225	0.0625	0.0341	-0.0057	0.0726	1.9417	0.3169	0.2504	-0.0518	0.3538
Median	0.1841	-0.0693	-0.0804	-0.1226	-0.0658	1.0652	-0.4310	-0.6711	-1.0085	-0.4173
25%	0.0223	-0.1846	-0.1869	-0.2506	-0.1765	0.1720	-1.2333	-1.5738	-1.8517	-1.2953
10%	-0.1029	-0.3325	-0.3346	-0.4031	-0.3335	-0.8232	-2.0558	-2.4835	-2.7864	-2.1905
5%	-0.1909	-0.4525	-0.4447	-0.5280	-0.4499	-1.4254	-2.5491	-2.9845	-3.3156	-2.7209
1%	-0.4353	-0.7310	-0.7757	-0.8856	-0.8025	-2.6052	-3.7717	-4.5878	-4.6634	-4.0465

	Panel B: Performance Evaluation Disagreement									
Mean Diff		0.3074	0.3253	0.3742	0.2883	% $\alpha_\varphi < \bar{\alpha}_{BC}$		99.17	97.67	93.32	96.98
(t-stat)		(6.728)	(6.070)	(5.276)	(5.422)	% $\alpha_\varphi sig < \bar{\alpha}_{BC}$		54.20	49.53	52.87	38.73
SD Diff		0.0603	0.0857	0.0823	0.0316	% $\alpha_\varphi > \bar{\alpha}_{BC}$		0.83	2.33	6.68	3.02
(t-stat)		(16.347)	(20.983)	(16.246)	(7.253)	% $\alpha_\varphi sig > \bar{\alpha}_{BC}$		0.00	0.00	0.04	0.00
Corr		0.7631	0.6549	0.3673	0.6903						

Table 3.5: Performance Disagreement for Power Utility Models

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the best clientele (BC) performance measure and the power utility models (POWER) with $\gamma = \{2, 4, 6\}$ and β estimated so that so that the resulting SDFs correctly price the risk free return. The BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between BC alphas and POWER alphas (the candidate alphas). On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$) and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlations.

	Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds								
	Performance				t-statistics				
	BC	POWER			BC	POWER			
		$\gamma = 2$	$\gamma = 4$	$\gamma = 6$		$\gamma = 2$	$\gamma = 4$	$\gamma = 6$	
Mean	0.2360	0.4293	0.4233	0.4174	1.0220	1.1739	1.1593	1.1447	
Std Dev	0.3344	0.2955	0.2958	0.2962	1.4333	0.7541	0.7540	0.7538	
(t-stat)	(3.351)	(6.890)	(6.796)	(6.692)					
99%	1.2708	1.0024	0.9914	0.9847	4.4955	2.7958	2.7804	2.7648	
95%	0.7851	0.8239	0.8178	0.8113	3.3544	2.3545	2.3374	2.3225	
90%	0.644	0.7350	0.7273	0.7203	2.7535	2.0951	2.0821	2.0660	
75%	0.4225	0.5896	0.5840	0.5776	1.9417	1.6734	1.6595	1.6462	
Median	0.1841	0.4441	0.4387	0.4331	1.0652	1.1845	1.1692	1.1564	
25%	0.0223	0.3014	0.2970	0.2924	0.1720	0.7374	0.7227	0.7081	
10%	-0.1029	0.1229	0.1169	0.1088	-0.8232	0.2279	0.2138	0.1979	
5%	-0.1909	-0.0287	-0.0320	-0.0375	-1.4254	-0.0536	-0.0747	-0.0905	
1%	-0.4353	-0.4253	-0.4373	-0.4493	-2.6052	-0.8816	-0.8964	-0.9113	

	Panel B: Performance Evaluation Disagreement							
Mean Diff	-0.1933	-0.1873	-0.1814		% $\alpha_\varphi < \bar{\alpha}_{BC}$	17.80	18.31	18.92
(t-stat)	(-3.357)	(-3.245)	(-3.134)		% $\alpha_\varphi sig < \bar{\alpha}_{BC}$	1.54	1.54	1.54
SD Diff	0.0389	0.0386	0.0382		% $\alpha_\varphi > \bar{\alpha}_{BC}$	82.20	81.69	81.08
(t-stat)	(8.424)	(8.339)	(8.230)		% $\alpha_\varphi sig > \bar{\alpha}_{BC}$	1.26	1.22	1.18
Corr	0.6297	0.6279	0.6260					

Table 3.6: Performance Disagreement for Habit-Formation Preference Models

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the best clientele (BC) performance measure and the habit-formation preference models (HABIT) with $\gamma = \{2, 4, 6\}$, $\mu = \{0.8, 0.9\}$ and β estimated so that so that the resulting SDFs correctly price the risk free return. The BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between BC alphas and HABIT alphas (the candidate alphas). On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$) and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlations.

Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds

| | Performance | | | | | | | t-statistics | | | | | | |
| | BC | HABIT $\gamma=2$ | | HABIT $\gamma=4$ | | HABIT $\gamma=6$ | | BC | HABIT $\gamma=2$ | | HABIT $\gamma=4$ | | HABIT $\gamma=6$ | |
		$\mu=0.8$	$\mu=0.9$	$\mu=0.8$	$\mu=0.9$	$\mu=0.8$	$\mu=0.9$		$\mu=0.8$	$\mu=0.9$	$\mu=0.8$	$\mu=0.9$	$\mu=0.8$	$\mu=0.9$
Mean	0.2360	0.4481	0.4670	0.4613	0.5016	0.4751	0.5422	1.0220	1.2075	1.2378	1.2245	1.2785	1.2397	1.3111
Std Dev	0.3344	0.2935	0.2938	0.2935	0.3015	0.2955	0.3209	1.4333	0.7380	0.7223	0.7221	0.6920	0.7066	0.6645
(t-stat)	(3.351)	(7.250)	(7.548)	(7.463)	(7.901)	(7.636)	(8.024)							
99%	1.2708	1.0369	1.0764	1.0675	1.1575	1.0955	1.2991	4.4955	2.7896	2.7866	2.7736	2.7963	2.7687	2.7962
95%	0.7851	0.8427	0.8643	0.8590	0.9290	0.8760	1.0156	3.3544	2.3576	2.3664	2.3513	2.3305	2.3312	2.3156
90%	0.644	0.7546	0.7735	0.7681	0.8206	0.7859	0.8854	2.7535	2.1001	2.1090	2.0958	2.0954	2.0842	2.0724
75%	0.4225	0.6047	0.6191	0.6134	0.6550	0.6253	0.6958	1.9417	1.6955	1.7073	1.6946	1.7246	1.7015	1.7207
Median	0.1841	0.4607	0.4755	0.4714	0.5005	0.4825	0.5279	1.0652	1.2243	1.2554	1.2416	1.2924	1.2522	1.3306
25%	0.0223	0.3252	0.3399	0.3355	0.3668	0.3463	0.3921	0.1720	0.7815	0.8290	0.8174	0.8903	0.8409	0.9560
10%	-0.1029	0.1528	0.1804	0.1763	0.2155	0.1908	0.2529	-0.8232	0.3035	0.3634	0.3496	0.4711	0.3952	0.5773
5%	-0.1909	0.0130	0.0424	0.0392	0.0884	0.0553	0.1445	-1.4254	0.0246	0.0938	0.0756	0.1875	0.1122	0.2898
1%	-0.4353	-0.3891	-0.3407	-0.3534	-0.2859	-0.3096	-0.2613	-2.6052	-0.8304	-0.7996	-0.8132	-0.7396	-0.7810	-0.7631

Panel B: Performance Evaluation Disagreement

	BC	$\gamma=2$ $\mu=0.8$	$\gamma=2$ $\mu=0.9$	$\gamma=4$ $\mu=0.8$	$\gamma=4$ $\mu=0.9$	$\gamma=6$ $\mu=0.8$	$\gamma=6$ $\mu=0.9$		$\gamma=2$ $\mu=0.8$	$\gamma=2$ $\mu=0.9$	$\gamma=4$ $\mu=0.8$	$\gamma=4$ $\mu=0.9$	$\gamma=6$ $\mu=0.8$	$\gamma=6$ $\mu=0.9$
Mean Diff	-0.2121	-0.2310	-0.2253	-0.2656	-0.2391	-0.3062		% $\alpha_\varphi < \bar{\alpha}_{BC}$	15.87	13.89	14.43	11.27	12.99	8.79
(t-stat)	(-3.716)	(-4.058)	(-3.953)	(-4.595)	(-4.181)	(-5.056)		% $\alpha_\varphi sig < \bar{\alpha}_{BC}$	1.54	1.54	1.54	1.54	1.51	1.54
SD Diff	0.0409	0.0406	0.0409	0.0329	0.0389	0.0135		% $\alpha_\varphi > \bar{\alpha}_{BC}$	84.13	86.11	85.57	88.73	87.01	91.21
(t-stat)	(8.933)	(8.883)	(8.943)	(7.064)	(8.467)	(2.761)		% $\alpha_\varphi sig > \bar{\alpha}_{BC}$	1.36	1.62	1.51	1.94	1.58	2.44
Corr	0.6343	0.6365	0.6354	0.6317	0.6345	0.6156								

Table 3.7: Performance Disagreement for the Manipulation Proof Performance Measure

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the best clientele (BC) performance measure and the manipulation proof performance measure (MPPM). The BC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between BC alphas and MPPM alphas (the candidate alphas). On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \bar{\alpha}_{BC}$), significantly smaller (% $\alpha_\varphi sig < \bar{\alpha}_{BC}$), larger (% $\alpha_\varphi > \bar{\alpha}_{BC}$) and significantly larger (% $\alpha_\varphi sig > \bar{\alpha}_{BC}$) than the BC alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlations.

Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds

| | | Performance | | | | t-statistics | | |
| | BC | MPPM | | | BC | MPPM | | |
		$\gamma = 2$	$\gamma = 3$	$\gamma = 4$		2	3	4
Mean	0.2360	0.1205	-0.0496	-0.2283	1.0220	0.4514	0.0896	-0.2582
Std Dev	0.3344	0.4077	0.6492	1.0347	1.4333	0.7192	0.7209	0.7315
(t-stat)	(3.351)	(1.403)	(-0.363)	(-1.048)				
99%	1.2708	0.6482	0.5374	0.4405	4.4955	2.0752	1.7383	1.4441
95%	0.7851	0.5152	0.3847	0.2772	3.3544	1.5838	1.2380	0.8980
90%	0.644	0.4351	0.3121	0.2018	2.7535	1.3297	0.9662	0.6066
75%	0.4225	0.3160	0.1876	0.0663	1.9417	0.9003	0.5360	0.1762
Median	0.1841	0.1813	0.0384	-0.1015	1.0652	0.4603	0.0916	-0.2480
25%	0.0223	0.0143	-0.1440	-0.3320	0.1720	0.0354	-0.3241	-0.6785
10%	-0.1029	-0.2301	-0.4451	-0.6879	-0.8232	-0.4428	-0.7792	-1.1311
5%	-0.1909	-0.4793	-0.7356	-1.0786	-1.4254	-0.7147	-1.0747	-1.4422
1%	-0.4353	-1.1260	-1.6497	-2.3417	-2.6052	-1.3921	-1.7907	-2.2374

Panel B: Performance Evaluation Disagreement

	Mean Diff	0.1155	0.2855	0.4642	% $\alpha_\varphi < \bar{\alpha}_{BC}$	53.19	69.78	84.03
	(t-stat)	(1.236)	(2.028)	(2.125)	% $\alpha_\varphi sig < \bar{\alpha}_{BC}$	2.40	4.63	9.94
	SD Diff	-0.0733	-0.3148	-0.7003	% $\alpha_\varphi > \bar{\alpha}_{BC}$	46.81	30.22	15.97
	(t-stat)	(-11.033)	(-38.402)	(-74.002)	% $\alpha_\varphi sig > \bar{\alpha}_{BC}$	0.11	0.00	0.00
	Corr	0.2985	0.1980	0.1530				

132

Table 3.8: Performance Disagreement from the Worst Clientele Performance Alphas

This table shows statistics on the cross-sectional distribution of monthly SDF alphas estimated with the worst clientele (WC) performance measure and five candidate models, i.e., the FF and CARHART unconditional linear factor models, the CFF and CCARHART conditional linear factor models and the manipulation proof performance measure with a risk aversion coefficient of $\gamma = 4$ (MPPM(4)). The WC measure allows for a maximum Sharpe ratio of $h^* + 0.5hMKT$ (see definition in section 3.3.2) and uses ten industry portfolios and the risk-free asset as passive portfolios. Panel A provides the mean, standard deviation (Std Dev) and selected percentiles of the distributions of estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of the estimated alphas (see test description in section 3.3.3). Panel B gives results on the disagreement between WC alphas and candidate model alphas. On the left side of the panel, it reports differences in mean alphas and standard deviations of alphas, their t-statistics (t-stat), and correlations between alphas (Corr). On the right side of the panel, it reports proportions of estimated candidate alphas that are smaller (% $\alpha_\varphi < \underline{\alpha}_{WC}$), significantly smaller (% $\alpha_\varphi sig < \underline{\alpha}_{WC}$), larger (% $\alpha_\varphi > \underline{\alpha}_{WC}$) and significantly larger (% $\alpha_\varphi sig > \underline{\alpha}_{WC}$) than the worst clientele alphas. The data (see description in table 3.1) are from January 1984 to December 2012. All statistics are in percentage except the t-statistics and the correlations.

Panel A: Alphas and t-statistics for the Cross-Section of Mutual Funds											
	Performance						t-statistics					
	WC	FF	CARHART	CFF	CCARHART	MPPM(4)	WC	FF	CARHART	CFF	CCARHART	MPPM(4)
Mean	-0.5938	-0.0866	-0.1200	-0.0893	-0.1382	-0.2283	-3.5688	-0.5515	-0.7398	-0.6974	-0.9863	-0.2582
Std Dev	0.3883	0.2576	0.2456	0.2487	0.2521	1.0347	1.4261	1.3061	1.2534	1.4016	1.4104	0.7315
(t-stat)	(-7.261)	(-1.597)	(-2.321)	(-1.705)	(-2.603)	(-1.048)						
99%	-0.0667	0.4354	0.4231	0.4640	0.4677	0.4405	-0.5195	2.3957	2.2289	2.3854	2.1929	1.4441
95%	-0.1978	0.2376	0.2120	0.2423	0.1955	0.2772	-1.3712	1.4457	1.3169	1.4461	1.3115	0.8980
90%	-0.2572	0.1643	0.1271	0.1681	0.1166	0.2018	-1.8549	1.0263	0.8321	1.0373	0.8311	0.6066
75%	-0.3625	0.0431	0.0109	0.0341	-0.0057	0.0663	-2.5970	0.3046	0.0731	0.2504	-0.0518	0.1762
Median	-0.5140	-0.0654	-0.1016	-0.0804	-0.1226	-0.1015	-3.4628	-0.5098	-0.7466	-0.6711	-1.0085	-0.2480
25%	-0.7159	-0.1847	-0.2214	-0.1869	-0.2506	-0.3320	-4.4925	-1.3427	-1.5149	-1.5738	-1.8517	-0.6785
10%	-1.0216	-0.3376	-0.3811	-0.3346	-0.4031	-0.6879	-5.3940	-2.2267	-2.2580	-2.4835	-2.7864	-1.1311
5%	-1.2511	-0.4622	-0.5037	-0.4447	-0.5280	-1.0786	-6.0399	-2.6916	-2.7996	-2.9845	-3.3156	-1.4422
1%	-1.8508	-0.8416	-0.8329	-0.7757	-0.8856	-2.3417	-7.2135	-4.1733	-4.1538	-4.5878	-4.6634	-2.2374
Panel B: Performance Evaluation Disagreement											
Mean Difference		-0.5072	-0.4738	-0.5045	-0.4556	-0.3655	% $\alpha_\varphi < \underline{\alpha}_{WC}$	0.00	0.83	0.04	0.65	5.31
(t-stat)		(-9.269)	(-7.711)	(-7.757)	(-6.850)	(-2.107)	% $\alpha_\varphi sig < \underline{\alpha}_{WC}$	0.00	0.00	0.00	0.00	0.11
SD Difference		0.1307	0.1427	0.1396	0.1362	-0.6464	% $\alpha_\varphi > \underline{\alpha}_{WC}$	100.00	99.17	99.96	99.35	94.69
(t-stat)		(33.560)	(33.342)	(30.587)	(29.005)	(-81.946)	% $\alpha_\varphi sig > \underline{\alpha}_{WC}$	94.72	81.01	91.56	78.75	11.52
Corr		0.7479	0.6605	0.6071	0.5853	0.6756						

Figure 3.1: Standard Deviation of Stochastic Discount Factors for Performance Evaluation
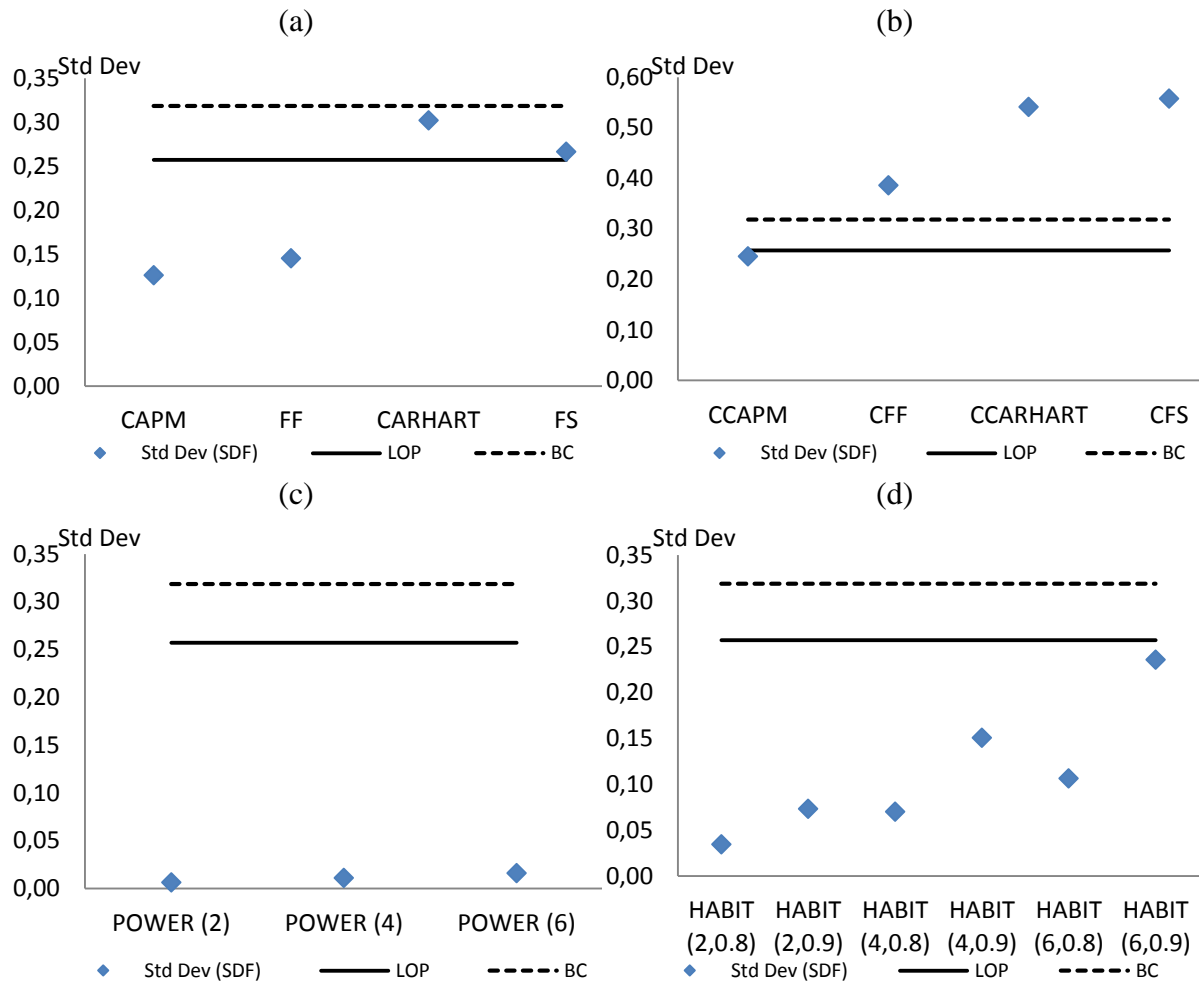
Notes: Figure 3.1 illustrates the standard deviation (Std Dev) of the stochastic discount factors (SDFs) for unconditional linear factor models (figure 3.1a), conditional linear factor models (figure 3.1b), power utility models (figure 3.1c) and habit-formation preference models (figure 3.1d). The continuous line is the standard deviation of the law-of-one-price (LOP) SDF. The dashed line is the standard deviation of the best clientele (BC) SDF.

4   Mutual Fund Styles and Clientele-Specific Performance Evaluation

Abstract

This paper develops clientele-specific performance measures based on the style preferences of mutual fund investors. Using a new approach that considers investor disagreement and better exploits style classification data, we investigate eight measures to represent clienteles with favorable preferences for size and value equity styles. Empirically, we find that the performance of funds assigned to styles associated with clienteles becomes neutral or positive when they are evaluated with their appropriate measure. The performance of the other funds is sensitive to the clienteles, and in particular their behavioral characteristics. The sign of the value added by the industry is ambiguous and depends on the choice of measures. Our findings support a significant role for style clienteles in performance evaluation.

Résumé

Cet article développe des mesures de performance spécifiques aux clientèles basées sur les préférences de style des investisseurs dans les fonds mutuels. Utilisant une nouvelle approche qui considère le désaccord entre investisseurs et exploite mieux les données de classification en styles, nous investiguons huit mesures pour représenter des clientèles avec des préférences favorables aux styles d'actions basés sur la taille et la valeur. Empiriquement, nous trouvons que la performance des fonds classés dans des styles associés à des clientèles devient neutre ou positive quand ceux-ci sont évalués avec leur mesure appropriée. La performance des autres fonds est sensible aux clientèles et, en particulier, à leurs caractéristiques comportementales. Le signe de la valeur ajoutée de l'industrie des fonds est ambigu et dépend du choix de mesure. Nos résultats supportent un rôle significatif des clientèles de style en évaluation de performance.

136

4.1 Introduction

Since the seminal contributions of Fama and French (1992, 1993), size (small-cap versus large-cap) and value (value versus growth) investment styles have grown so much in popularity among investors that equity style investing has become dominant in industry practices. Thousands of mutual funds now advertised themselves according to their size and value focuses, oftentimes starting with their names. They cater to and attract size and value clienteles who can rely on numerous style classification tools for their investment decisions. This economically significant style differentiation among funds and investors suggests that mutual fund performance evaluation should also differ for size and value clienteles.

Reviewing recent research, Ferson (2010) emphasizes the importance of identifying meaningful investor clienteles and developing clientele-specific performance measures to properly evaluate mutual funds. Many studies argue that investor heterogeneity and clientele effects influence the flow-performance relationship (Del Guercio and Tkac, 2002, Christoffersen, Evans and Musto, 2013), and are related to behavioral biases (Barber, Odean and Zheng, 2005, Bailey, Kumar and Ng, 2011), demographics and investor sophistication (Malloy and Zhu, 2004, Bailey, Kumar and Ng, 2011), investor monitoring and investment advice (James and Karceski, 2006, Bergstresser, Chalmers and Tufano, 2009, Del Guercio and Reuter, 2014), liquidity demands (Nanda, Narayanan and Warther, 2000), recession and risk aversion (Goetzmann and Massa, 2002, Glode, 2011), and taxation (Ivković and Weisbenner, 2009).

Given the predominance of style investing, equity styles are also relevant to identify meaningful investor clienteles. Barberis and Shleifer (2003) point out the growing interest of financial service firms to understand style preferences and there is a related literature that attempts to classify funds according to their correct styles (see Kim, Shukla and Tomas (2000), Brown and Goetzmann (1997) and Dibartolomeo and Witkowski (1997), among others). Also, Shefrin and Statman (1995, 2003), Blackburn, Goetzmann and Ukhov (2013) and Shefrin (2015) study the judgments, sentiment sensitivity and trading behavior of style

investors. They find that value and small-cap investors tend to be pessimists and contrarians, while growth and large-cap investors tend to be optimists and trend followers.

Despite these contributions, the growing literature on style investors and clientele effects in mutual funds has not focused on developing performance measures that account for the heterogeneous preferences of the various clienteles. This paper develops clientele-specific performance measures based on the implied style preferences of mutual fund investors and empirically investigates whether performance evaluation differs for size and value clienteles. Our approach has two distinctive features.

First, we develop a performance framework with investor disagreement that allows for the identification of meaningful stochastic discount factors (SDFs) for style clienteles. The framework is based on the SDF alpha approach of Chen and Knez (1996) and is a refinement of the best clientele approach of Chrétien and Kammoun (2015). The performance measures are extracted from a set of SDFs admissible under the law-of-one-price and no-good-deal conditions, an incomplete market setup initiated by Cochrane and Saá-Requejo (2000). Our identification strategy uses the empirical results of Ferson and Lin (2014) on the magnitude of investor disagreement and assumes that style portfolios are representative investments for style clienteles. This framework ensures that the performance measures correctly price passive portfolios and generates a realistic and sufficiently large disagreement to differentiate the style clienteles.

Second, we introduce a new method to better exploit available style classification data in the CRSP mutual fund database. The method focuses on Lipper objective codes as they cover the most relevant period and facilitate the classification of funds into size and value styles. It also accounts for code changes and missing codes by using a threshold for style inclusion that considers whether a fund has been assigned to a code for enough time. The method is thus based on publicly available information from a leading industry provider that clienteles could presumably consult to form their investment decisions, considers the stability and quality of the code data, and avoid the ad hoc attribution of codes into style categories. Using the method, we classify the funds and their clienteles into eight size and value categories: four general styles attractive to broader clienteles (small-

cap, large-cap, value and growth) and four specific styles attractive to more specialized clienteles (large-cap growth, small-cap growth, large-cap value and small-cap value).

Using a sample of 2530 actively-managed open-ended U.S. equity mutual funds with monthly returns from 1998 to 2012, our empirical investigation makes three contributions. First, we study the economic properties of the SDFs identified by our approach with a nonlinear approximation that can be interpreted from a rational or behavioral perspective, following Dittmar (2002) and Shefrin (2008, 2009). We find that the preferences implied by the style clientele SDFs have similar risk aversion but differ in their behavioral features. The SDFs imply value and small-cap investors who tend to be pessimists and contrarians, and growth and large-cap investors who tend to be optimists and trend followers. Hence, the SDFs are not only different enough to generate performance disagreement, but they are also consistent with the individual investor behavior documented by Shefrin and Statman (1995, 2003), Blackburn, Goetzmann and Ukhov (2013) and Shefrin (2015).

Second, we implement a clientele-specific performance evaluation using the style clientele SDFs. We find that the funds assigned to size and value styles have neutral to positive average alphas when they are evaluated with their appropriate clientele-specific measure. Hence, the evaluation considers relevant clienteles, fund performance is more positive than existing evidence shows (see Fama and French (2010) and Barras, Scaillet and Wermers (2010) for recent examples). The performance of the other funds is sensitive to the clienteles. Specifically, their average alphas are significantly negative to neutral for value and small-cap clienteles, but neutral to significantly positive for growth and large-cap clienteles. For these funds, the behavioral features of the investor SDFs are important determinants of performance evaluation.

Third, we document the value added of the mutual fund industry from the perspective of different style clienteles. Given that funds are attributed multiple performance values by the SDFs, we examine many cross-sectional performance distributions by considering either the minimum or maximum alpha of each fund. The results show that the sign of the value added is ambiguous and depends on the choice of

measures. But they suggest that the value added is greater for growth and large-cap investors than for value and small-cap investors, although the difference is sensitive to the magnitude of disagreement allowed in our performance framework.

Overall, this paper shows that preferences and performance evaluations differ for size and value mutual fund clienteles. It is an important first step toward confirming the conjecture of Ferson (2010) that clientele-specific measures based on meaningful investor clienteles might be necessary to properly evaluate mutual funds. The clientele-specific performance evaluation approach we propose can also serve as a useful framework for developing measures that account for the clientele effects documented in the literature. Such measures could be important because our findings support a significant role for clienteles in performance evaluation.

The remainder of this paper is organized as follows. Section 4.2 develops a theoretical framework for performance measurement with style clienteles. Section 4.3 presents the methodology for estimating and summarizing the results. Section 4.4 describes the data. Section 4.5 presents the style classification method and results. Section 4.6 reports and analyzes the empirical results. Finally, section 4.7 provides the conclusion.

4.2 Performance Measures for Style Clienteles: Theoretical Framework

Using the stochastic discount factor (hereafter SDF) approach, this section develops a framework for performance evaluation under the assumption that there are style clienteles. First, we present a basic setup for performance evaluation with investor disagreement. Second, we describe our strategy for identifying meaningful SDFs for style clienteles. Third, we propose style-clientele-specific performance measures for the evaluation of individual mutual funds.

4.2.1 Performance Evaluation with Investor Disagreement

This paper exploits the general framework of the SDF performance approach, first proposed by Glosten and Jagannathan (1994) and Chen and Knez (1996), to allow for investor disagreement that occurs when an investor evaluates a fund differently from another investor (see Ahn, Cao and Chrétien (2009), Ferson (2010) and Ferson and Lin

140

(2014) and Chrétien and Kammoun (2015)). In this approach, the performance, or SDF alpha, is defined from the following equation:

(1)
$$\alpha_{MF} = E[m\, R_{MF}] - 1,$$

where m is the SDF of an investor interested in valuing an individual mutual fund with gross return R_{MF}. Unlike most of the literature, this paper does not assume that there is a unique SDF for all investors. Instead, we view a fund investor's SDF as part of a set M of admissible SDFs that allow for investor disagreement and heterogeneous preferences.

Under general conditions, in an incomplete market, Chen and Knez (1996) show that there is an infinite number of admissible SDFs that generates a potentially *infinite alpha disagreement* between investors. To restrict SDFs in an economically meaningful way, we follow Hansen and Jagannathan (1991), Cochrane and Saá-Requejo (2000) and Chrétien and Kammoun (2015), and impose two conditions: the law-of-one-price condition and the no-good-deal condition. The law-of-one-price condition assumes that mutual fund investors give zero performance to passive portfolios. The no-good-deal condition assumes that investors do not allow investment opportunities that have too large Sharpe ratios.

These restrictions are equivalent to limiting the variability of admissible SDFs. Specifically, Cochrane and Saá-Requejo (2000) show that, under these two conditions, for all $m \in M$, the SDF standard deviation is bounded as follows:

(2)
$$\frac{h^*}{R_F} \leq \sigma(m) \leq \frac{\bar{h}}{R_F},$$

where h^* is the optimal Sharpe ratio attainable from the passive portfolios, $\bar{h}$ is the maximum allowable Sharpe ratio for investments not considered good deals and R_F is the gross risk-free return. Intuitively, we can interpret the SDFs as representing the marginal preferences of mutual fund investors. Equation (2) stipulates that the variability of their marginal utilities should be large enough to correctly price existing passive portfolios, but

small enough to rule out implausibly high risk aversion that would allow good deals to be viable.

The SDFs admissible under these conditions are advantageous for performance evaluation for two reasons. First, they do not suffer from the benchmark choice problem that occurs when a selected benchmark does not correctly price passive portfolios (Chen and Knez, 1996, Fama, 1998, Ahn, Cao and Chrétien, 2009, and Chrétien and Kammoun, 2015). Second, they allow for a reasonable investor disagreement consistent with the analysis of Ferson and Lin (2014). Specifically, as the set M of SDFs is closed and convex, there is a *finite alpha disagreement* between investors because it is possible to find lower and upper performance bounds:

$$(3) \qquad \underline{\alpha}_{MF} \leq \alpha_{MF} \leq \bar{\alpha}_{MF},$$

Chrétien and Kammoun (2015) call these bounds the worst and best clientele alphas, respectively. Using equity mutual funds, they document disagreement values comparable to those of Ferson and Lin (2014) and significant enough to change the value of funds from negative to positive, depending on the clienteles. Their results confirm the analysis of Ferson and Lin (2014) on the economic importance of investor disagreement and clientele effects in performance evaluation.

Although Chrétien and Kammoun (2015) show the relevancy of the admissible SDFs used in this study, the goals of the papers are different. By estimating fund-specific bounds, they document the performance for the best and worst clienteles of each individual mutual fund. Instead of focusing on fund-specific clienteles, this paper examines general clienteles interested in particular investment styles within the mutual fund industry. The next subsection provides a strategy for selecting specific SDFs in M that identify these general style clienteles.

4.2.2   Stochastic Discount Factors for Style Clienteles

Ferson (2010) emphasizes the importance of identifying meaningful investor clienteles in mutual funds. This paper assumes that grouping funds by their investment style is a

relevant strategy for this purpose. Let R_s be the gross return on a portfolio of funds grouped according to their similar style s. For simplicity, let S represents the set of styles under consideration (i.e., $S = \{\text{Value, Growth, etc.}\}$) as well as the total number of styles. Using any $m \in M$, we can measure the performance, or alpha, of this style portfolio by:

(4) $$\alpha_s = E[m\,R_s] - 1.$$

Given the existence of many admissible SDFs, investors likely disagree on the alpha of this style portfolio. Our approach stipulates that the SDF at the upper performance bound is a meaningful SDF for representing the clienteles favorable to the style.

Specifically, to find the SDF for clienteles most favorable to the investment style providing a return R_s, we solve the following problem:

(5) $$\bar{\alpha}_s = \sup_{m \in M} E[m\,R_s] - 1,$$

(6) $$\text{subject to } E[m\,\mathbf{R_K}] = \mathbf{1}, \ E[m^2] \leq \frac{(1+\bar{h}^2)}{R_F{}^2},$$

where $E[m\,\mathbf{R_K}] = \mathbf{1}$ is law-of-one-price condition, with $\mathbf{R_K}$ being a vector of gross returns on K passive portfolios and $\mathbf{1}$ is a $K \times 1$ unit vector, and $E[m^2] \leq \frac{(1+\bar{h}^2)}{R_F{}^2}$ is the no-good-deal condition, with $\bar{h}$ being the maximum Sharpe ratio allowed. Cochrane and Saá-Requejo (2000) demonstrate that this problem has the following closed-form solution:

(7) $$\bar{\alpha}_s = E[\bar{m}_s\,R_s] - 1,$$

with:

(8) $$\bar{m}_s = m_{LOP} + v_s w_s,$$

(9)
$$m_{LOP} = \mathbf{a}'\mathbf{R_K},$$

(10)
$$w_s = R_s - \mathbf{c_s}'\mathbf{R_K},$$

where:

(11)
$$\mathbf{a}' = \mathbf{1}'E[\mathbf{R_K}\,\mathbf{R_K'}]^{-1},$$

(12)
$$\mathbf{c_s}' = E[R_s\,\mathbf{R_K'}]\,E[\mathbf{R_K}\,\mathbf{R_K'}]^{-1},$$

(13)
$$v_s = \sqrt{\frac{\left(\frac{(1+\bar{h}^2)}{R_F^{\,2}} - E[m_{LOP}^2]\right)}{E[w_s^2]}}.$$

In this solution, $\bar{\alpha}_s$ is the upper bound on the (expected) alpha of a style portfolio, i.e., the highest average performance value found from the heterogeneous investors with SDFs in the set M. The SDF that solves the problem is denoted by $\bar{m}_s$ and called the "style clientele" SDF for investment style s. It identifies the SDF for the class of investors most favorable to style s and hence provides a way to obtain meaningful SDFs for various style clienteles. Its first component, m_{LOP}, is shown by Hansen and Jagannathan (1991) to be the minimum volatility SDF under the law-of-one-price condition, and used by Chen and Knez (1996) for their LOP performance measure. As indicated by the notation, it is common to all style clientele SDFs because it depends only on the passive portfolio returns and does not vary with the style portfolio investigated.

Its second component, $v_s w_s$, is a clientele-specific component that accounts for investor disagreement. The replicating error term w_s is the difference between the style portfolio return and the best replicating payoff constructed from the passive portfolio returns. The disagreement parameter v_s accounts for the no-good-deal restriction and is a function of the maximum Sharpe ratio $\bar{h}$. The second component thus indicates that

144

investors mostly agree on the evaluation of style portfolios that have easy-to-replicate returns or when they consider most allowable investment opportunities as good deals. In other instances, there can be significant clientele effects in performance evaluation.

4.2.3 Style-Clientele-Specific Performance Evaluation

Ferson (2010) discusses the importance of clientele-specific performance measures to properly evaluate mutual funds and calls for the development of such measures. Section 4.2.2 provides a strategy to identify meaningful SDFs for style clienteles that can be exploited for such a purpose. Assuming that S representative investment style portfolios can be formed, the strategy yields S different style clientele SDFs, i.e., $\bar{m}_s$ for all $s \in S$. This section describes how these SDFs can be used in a style-clientele-specific performance evaluation.

The evaluation necessitates that we distinguish between two types of individual mutual funds. The first type includes funds assigned to investment styles associated with clienteles. Performance measurement for these funds is relatively straightforward because we can use their associated style clientele SDFs. Let $R_{MF,s}$ be the gross return on a fund assigned to style s associated with a clientele. Its unique style-clientele specific alpha is given by:

$$(14) \qquad\qquad \alpha_{MF,s} = E\big[\bar{m}_s R_{MF,s}\big] - 1.$$

The second type includes funds not assigned to investment styles associated with clienteles. Such funds could have a style different from the ones considered, frequent changes of styles through time or missing style classification information. Because it is not possible to select a unique style clientele SDF for these funds, we instead examine their performance for all clienteles. Let $R_{MF,-}$ be the gross return on a fund not assigned to a style associated with a clientele. Its multiple style-clientele specific alphas are given by:

$$(15) \qquad\qquad \alpha_{MF,s} = E\big[\bar{m}_s R_{MF,-}\big] - 1, \text{ for all } s \in S.$$

Equations (14) and (15) simply specify that we evaluate the performance of individual mutual funds using their appropriate style-clientele performance measures. When a fund can be related to a clear style, it likely caters to and attracts the clienteles most favorable to its style. It thus makes sense to use the style clientele SDF associated with its style for performance measurement. Even though this SDF implies preferences that are favorable to the style of the fund, it can still evaluate the fund positively or negatively, depending on how the fund differs from the style portfolio that includes other funds of similar style. However, when a fund cannot be related to a clear style, it could presumably cater to and attract all types of clienteles. By considering all the style-clientele-specific performance measures for evaluation, we can account for the disagreement between clienteles on the value of this fund.

4.3   Methodology

4.3.1   Estimation

The style clientele SDF has an analytical solution that has $2K + 1$ parameters to estimate. We use the generalized method of moments (GMM) of Hansen (1982) for estimation and inferences. For a sample of T observations, we rely on a just-identified system with the following moments:

(16)
$$\frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}})\mathbf{R_{Kt}}] - \mathbf{1} = 0,$$

(17)
$$\frac{1}{T}\sum_{t=1}^{T}[(R_{st} - \mathbf{c}_s{}'\mathbf{R_{Kt}})\mathbf{R_{Kt}}] = 0,$$

(18)
$$\frac{1}{T}\sum_{t=1}^{T}[(\mathbf{a}'\mathbf{R_{Kt}} + v_s(R_{st} - \mathbf{c}_s{}'\mathbf{R_{Kt}}))R_{st}] - 1 - \bar{\alpha}_s = 0.$$

Equation (16) uses the fundamental SDF pricing equation to set the K moments needed to correctly price the K passive portfolios, estimating the $K \times 1$ vector of coefficients $\mathbf{a}$. It allows the estimation of the LOP SDF of Chen and Knez (1996), $m_{LOPt} = \mathbf{a}'\mathbf{R_{Kt}}$. Equation (17) specifies the K orthogonality conditions between the replication error term, $w_{st} = R_{st} - \mathbf{c_s}'\mathbf{R_{Kt}}$, and passive portfolio returns. They estimate the linear combination of passive portfolio returns that best replicates the style portfolio return, estimating the $K \times 1$ vector of coefficients $\mathbf{c}$. The moment in equation (18) allows the estimation of the disagreement parameter v_s, which completes the estimation of the style clientele SDF $\bar{m}_{st} = \mathbf{a}'\mathbf{R_{Kt}} + v_s(R_{st} - \mathbf{c_s}'\mathbf{R_{Kt}})$. The estimation of v_s implements an empirical version of equation (7) and requires an exogenously specified value for $\bar{\alpha}_s$, the upper performance bound on the style portfolio or expected alpha for clienteles most favorable to the style. Section 4.3.2 expands on this method for estimating v_s and discusses our approach to select the value of $\bar{\alpha}_s$, exploiting the results of Ferson and Lin (2014).

To evaluate the performance of individual mutual funds with the style clientele SDF, we add the following moment to the previous system:

(19)
$$\frac{1}{T}\sum_{t=1}^{T}[\bar{m}_{st}R_{MFt}] - 1 - \alpha_{MF,s} = 0,$$

where the gross mutual fund return R_{MFt} represents the return on a fund assigned to a style s associated with a clientele, $R_{MF,st}$, or the return on a fund not assigned to a style associated with a clientele, $R_{MF,-t}$. This moment represents the empirical counterpart of the style-clientele-specific performance measures proposed in this paper.

In all estimation cases, our procedure estimates alpha separately for each fund. Farnsworth, Ferson, Jackson and Todd (2002) demonstrate that estimating this system for one fund at a time produces the same point estimates and standard errors for alpha as a system that includes an arbitrary number of funds. Also, the parameter estimates are not influenced by the choice of the weighting matrix in GMM because the systems are just identified. Their statistical significance is assessed with the asymptotic properties of GMM

derived by Hansen (1982), with standard errors adjusted for conditional heteroskedasticity and serial correlation using the method of Newey and West (1987) with two lags.[1]

4.3.2 Methodological Choices

Empirically, this paper investigates the style clienteles and performance of actively-managed U.S. equity mutual funds. The estimation of the previous system requires three important methodological choices in relation to these funds. The first and perhaps most important choice is the style classification of funds and the formation of style portfolios. Section 4.5 examines this choice carefully and introduces a new approach to better exploit available style classification data in the *CRSP Survivor-Bias-Free US Mutual Fund Database*. It shows that the resulting style classifications and portfolios, defined by size and value sorts, are relevant for our methodology. This subsection discusses the second and third choices: the value of the upper performance bound on the style portfolio return, $\bar{\alpha}_s$, and the assets to form the passive portfolio returns $\mathbf{R_K}$.

As discussed in section 4.3.1, the estimation of v_s with equation (18) necessitates a pre-specified value of $\bar{\alpha}_s$. This value should be a realistic alpha for clienteles attracted to the style, because it ultimately identifies the SDF meant to capture the marginal preferences of the class of investors most favorable to the style. Ferson and Lin (2014) provide an indirect way to select a relevant value and we exploit their results. They study the effects of investor disagreement on performance evaluation and estimate bounds for the expected disagreement with a traditional alpha for various benchmark returns. For example, using the factors of Fama and French (1993) and a sample of funds similar to ours, they find mean and median monthly bound values of 0.248% and 0.212%, respectively (Ferson and Lin, 2014, table 3, panel A). Hence, their results suggest that clienteles most favorable to a fund could realistically expect mean and median alphas of $\alpha_{FF} + 0.248\%$ and $\alpha_{FF} + 0.212\%$, respectively, where α_{FF} is the Fama-French alpha.

---

[1] Chrétien and Kammoun (2015) investigate the finite sample properties of SDF alphas estimated with a similar system of moments and find that inferences are generally robust to finite sample issues. They also show that *t*-statistics for their alphas are similar when using no lag or four lags. The choice of two lags is meant to control for the small but significant serial correlation in the returns of some funds (that might invest in thinly traded assets).

148

In section 4.5.3, we use these results to fix the value of $\bar{\alpha}_s$. Specifically, we estimate the Fama-French alpha for each style portfolio under consideration and add the previously mentioned mean and median bound estimates to compute realistic values of alphas for favorable clienteles. Based on the range of values obtained, we then select an economically relevant value for $\bar{\alpha}_s$. Although this value represents our basic choice, we also empirically examine the effects on the results of other sensible values in section 4.6.4.

This method for estimating v_s differs from the method used by Cochrane and Saá-Requejo (2000) and Chrétien and Kammoun (2015). These papers estimate v_s by exploiting the no-good-deal restriction that the solution for the style clientele SDF must meet: $E[\bar{m}_{st}^2] = \frac{(1+\bar{h}^2)}{R_F^2}$. They rely on exogenous values for the maximum Sharpe ratio $\bar{h}$, selected with guidance from the literature, and the risk-free rate R_F. Chrétien and Kammoun (2015) find that their alpha estimates are relatively sensitive to the choice of $\bar{h}$, although their conclusions hold for all reasonable values investigated. The estimation method in this paper avoids the specification of a maximum Sharpe ratio and instead uses available evidence on performance disagreement to choose a relevant alpha for clienteles favorable to a style.

The third important methodological choice is the selection of passive portfolios. These portfolios are central to m_{LOP}, the part of $\bar{m}_s$ common to all style clienteles. In our setup, no investor disagrees on their value. We choose a risk-free asset and ten industry portfolios as passive portfolios and details on the specific return series are provided in the data section. The inclusion of the risk-free asset controls for cash positions in equity funds and follows the recommendation of Dahlquist and Söderland (1999) to fix the mean of the SDFs to a reasonable value. Industry portfolios are widely used as basis assets in empirical asset pricing and performance evaluation, and categorization by industry is a common practice for mutual fund investors and researchers. For the purpose of this study, selecting industry portfolios instead of size and value Fama-French portfolios also provides a crucial benefit: it facilitates the identification of style clientele SDFs that, because they generate disagreement on the value of the style portfolios, are more likely to meaningfully represent different style clienteles. It is thus useful to exclude portfolios formed on criteria related to

the investigated styles from the set of passive portfolios to avoid obtaining SDFs that lead to no disagreement on them.

Of course, other sets of passive portfolios could be relevant. Chrétien and Kammoun (2015) examine the sensitivity of their upper performance bound results to three sets of passive portfolios and find that their results are robust. Ferson and Lin (2014) also find that their results on investor disagreement are robust to changes of benchmark returns. These findings suggest that our results on disagreement between style clienteles should not be qualitatively affected by the choice of passive portfolios.

4.3.3 Cross-Sectional Performance Statistics

We summarize the empirical results by using numerous cross-sectional statistics. To examine the distribution of the alpha estimates for each style-clientele-specific performance measure, we show the mean, standard deviation and selected percentiles of the distributions of the estimated alphas and their corresponding t-statistics, computed as $t = \hat{\alpha}/\hat{\sigma}_{\hat{\alpha}}$, where $\hat{\alpha}$ is the estimated alpha and $\hat{\sigma}_{\hat{\alpha}}$ is its Newey-West standard error. We also provide t-statistics to test for the hypothesis that the cross-sectional mean of the estimated alphas is equal to zero. This test assumes that the distribution of the alphas across funds is multivariate normal with a mean of zero, a standard deviation equal to the observed cross-sectional standard deviation, and a correlation between any two alphas of 0.044. This value corresponds to the cross-sectional dependence in performance among equity funds, adjusted for data overlap, documented in Barras, Scaillet and Wermers (2010, p. 193) and Ferson and Chen (2015, appendix, p. 62). Finally, we present the proportions of estimated alphas that are positive and negative.

4.4 Data

4.4.1 Mutual Fund Returns

Our fund data consist of monthly returns on actively-managed open-ended U.S. equity mutual funds from January 1998 to December 2012. The data source is the *CRSP Survivor-Bias Free Mutual Fund US Database* and our starting date corresponds to the date of first availability of Lipper objectives codes, which are central to the style classification approach

introduced in section 4.5. Following Kacperczyk, Sialm and Zheng (2008), we exclude bond funds, balanced funds, money market funds, international funds, funds that are not strongly invested in common stocks, index funds and funds not opened to investors.[2]

Our sample selection mitigates numerous mutual fund database biases documented in the literature. Survivorship bias is treated in the CRSP database. Selection bias does not matter for our study period, as Elton, Gruber and Blake (2001) and Fama and French (2010) show its presence only before 1984. To deal with back-fill and incubation biases, we follow Elton, Gruber and Blake (2001), Kacperczyk, Sialm and Zheng (2008) and Evans (2010). We eliminate observations before the organization date of the funds, funds with no reported organization date and funds without a name, since they tend to correspond to incubated funds. We also exclude funds with total net assets inferior to $15 million in the first year of entering the database.

As a last sampling choice, following Barras, Scaillet and Wermers (2010) and others, we impose a minimum fund return requirement of 60 months. While this screen introduces a weak survivorship bias, it is useful to obtain reliable statistical estimates with GMM. Barras, Scaillet and Wermers (2010) and Chrétien and Kammoun (2015) find that their performance results are similar when using a 36-month requirement instead of a 60-month requirement. We obtain a final sample of 2530 actively-managed open-ended U.S. equity mutual funds.

4.4.2 Passive Portfolio Returns

The passive portfolio returns are the monthly returns on the one-month U.S. Treasury bills (taken from CRSP) and ten industry portfolios (taken from Kenneth R. French's website). The industries are consumer nondurables (NoDur), consumer durables (Durbl), manufacturing (Manuf), energy (Enrgy), high technology (HiTec), telecommunication (Telcm), shops (Shops), healthcare (Hlth), utilities (Utils) and other sectors (Others).

---

[2] Specifically, we identify US equity funds by policy codes CS and Lipper objective codes EIEI, EMN, LCCE, LCGE, LCVE, MATC, MATD, MATH, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE or SCVE, and keep the funds only if they hold between 80% and 105% in common stocks. We exclude index funds identified by the Lipper objective codes SP and SPSP or with a name that includes the word "index", and funds not opened to investors by consulting the variable "open to investors".

4.4.3 Summary Statistics

Table 4.1 presents summary statistics for the monthly returns of the mutual funds (panel A) and passive portfolios (panel B). Panel A also includes summary statistics for the Fama-French SDF alphas with their corresponding *t*-statistics.[3] In panel A, the average mutual fund return is 0.485% with a standard deviation of 0.309%. The monthly Sharpe ratios have a mean of 0.045 and a standard deviation of 0.055, with a range from -0.306 to 0.232. The Fama-French alpha estimates have a mean of -0.064% (*t*-stat. = -1.16) and a standard deviation of 0.267%. Approximately 60% of funds have negative performance values. At the 5% level, approximately 10% (2%) of funds have significantly negative (positive) alpha estimates. In panel B, industry portfolios have mean returns from 0.356% to 0.999% and standard deviations from 3.816% to 8.233%.

4.5 Style Classification

This section introduces a new method to better exploit available style classification data in the *CRSP Survivor-Bias-Free US Mutual Fund Database*. It also shows that the resulting style classifications and style portfolios, defined by size and value sorts, are relevant choices toward identifying meaningful clienteles in mutual funds.

For the purpose of this study, a good style classification method should meet two objectives. First, it should be based on publicly and easily available information that clienteles could presumably consult to form their investment decisions. This objective leads us to avoid using the statistical methods proposed by Sharpe (1992), Brown and Goetzmann (1997) and others, to instead rely on existing classifications from industry providers like Lipper and Morningstar. Second, the classification should be based on styles commonly accepted by mutual fund investors. The high impact research of Fama and French (1992, 1993) has established the importance of size and value in the cross-section of

---

[3] The Fama-French SDF is a linear function of the market, size (SMB) and value (HML) factors available on Kenneth R. French's website. For each mutual fund, we estimate jointly the parameters of the Fama-French SDF and the alpha by using GMM with a just identified system. Specifically, we estimate the four parameters of the Fama-French SDF by requiring the SDF to correctly price the one-month Treasury bill returns and the three Fama-French factors, and the alpha by using a moment similar to equation (19), but with the Fama-French SDF replacing the style clientele SDF.

equity returns. Since these seminal contributions, size (small-cap versus large-cap) and value (value versus growth) investment styles have become dominant in industry practices. Over the last 20 years, many equity funds have advertised themselves according to their size and value focuses, oftentimes starting with their names. They cater to and attract size and value investors, who can rely on classification tools like the style box popularized by Morningstar and illustrated in figure 1. To reach our style classification goals, the method detailed in the rest of this section thus uses existing industry classifications to split funds into size and value style categories.

4.5.1 Style Classification Data from the CRSP Mutual Fund Database

The *CRSP Survivor-Bias Free Mutual Fund US Database* provides style codes from three sources: Wiesenberger, Strategic Insight and Lipper.[4] Each source is only available for a part of the CRSP sample period: from 1962 to 1993 for Wiesenberger, from 1993 to 1998 for Strategic Insight and since 1998 for Lipper. This paper focuses exclusively on the period with Lipper data for two main reasons. First, Lipper codes cover most of the period during which size and value investment styles are prevalent in the industry. The Lipper period is likely the most relevant period to identify size and value clienteles. Second, Lipper codes facilitate the classification of equity funds into size and value styles because they are named according to these styles. In particular, they include large-cap growth funds (code LCGE), small-cap growth funds (code SCGE), large-cap value funds (code LCVE) and small-cap value funds (code SCVE). Hence, focusing on the Lipper period allows us to avoid the ad hoc attribution of Wiesenberger and Strategic Insight codes into the style categories considered in this paper. For example, Pástor and Stambaugh (2002) use all three sources of codes but subjectively assign various codes to their seven broad investment objectives.[5] Lipper, a Thomson Reuters company, is also a global leader in supplying mutual fund information and thus a source regularly consulted by fund clienteles.[6]

---

[4] Morningstar is another widely-used source for fund classification but their codes are not available in the CRSP mutual fund database.

[5] In an attempt to extend our sample period to 1984, we examine a style classification based on the Lipper, Wiesenberger and Strategic Insight codes, prioritized in that order. We find that the codes from the latter two sources are not useful because of the difficulty to assign them to our categories. In particular, they do not

An important issue arising from Lipper codes in the CRSP database is that, although a unique classification code is typically attributed to a fund for a given period (which can be a month up to its full sample period), it can frequently change through time or be missing. Panel A of table 4.2 reports statistics to assess the stability and quality of the Lipper classification codes available for the 2530 equity mutual funds in our sample. It provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the numbers of code changes per fund (Nb Changes), the numbers of different codes per fund (Nb Styles), the maximum proportions of monthly observations that a fund has the same code (% Max) and the proportions of monthly observations that a fund presents a missing code (% Missing).

The results show that frequent changes in codes are a bigger issue than missing codes. On average, funds have 3.77 code changes and 3.19 different codes from 1998 to 2012. Less than 1% of the funds have no code change. But more than 10% of the funds have at least six code changes and four different codes. One fund even experiences 11 different codes. Although code changes are frequent, many funds keep the same code for a large fraction of their observations. On average, funds spend 67.46% of their sample observations with the same code and one fourth of the funds have the same code for at least 82% of their observations. The fraction of funds with less than half (a quarter) of their observations with the same code is approximately equal to 20% (0.5%). Finally, the proportions of months with missing codes average 12.83%. Only 1% of the funds have no missing code, but less than 5% of the funds have more than 25.75% of missing codes. Our style classification method accounts for both code changes and missing codes.

4.5.2 Style Classification Method and Results

This section describes the style classification method developed in this paper and presents the classification results for the mutual funds. Figure 1 presents a mutual fund style box

identify more large-cap value funds and small-cap value funds, and only convincingly identify 20 new small-cap growth funds.

[6] Their website mentions that: "Lipper's benchmarking and classifications are widely recognized as the industry standard by asset managers, fund companies and financial intermediaries. Our reliable fund data, fund awards designations and ratings information provide valued insight to advisors, media and individual investors."

that illustrates the classification. The method proceeds in five steps. First, we start by identifying all funds with at least one of the four Lipper codes that jointly consider the size and value styles (i.e., Lipper codes LCGE, SCGE, LCVE and SCVE). Second, to deal with code changes and missing codes, we compute, for each identified fund, the proportion of monthly return observations associated with each code attached to the fund. One minus the sum of these "Lipper code proportions" give the fraction of observations with a missing code.

Third, if its corresponding Lipper code proportion is above a given threshold, we assign a fund to one of the following four specialized styles: large-cap growth (LCG), small-cap growth (SCG), large-cap value (LCV) and small-cap value (SCV). For example, if we select a threshold of 50%, then a fund is classified as LCV only if its Lipper code is LCVE for 50% of its return observations. This "threshold for style inclusion" ensures that the Lipper code information is sufficiently reliable for the fund to be categorized, i.e. a fund with different or missing codes for a significant fraction of its sample period is not assigned to a style. As LCG, SCG, LCV and SCV funds are in the corners of the style box in figure 1, we call them "corner" funds for brevity.

Fourth, we form four broader style categories by combining corner funds with a common style. Hence, we assign the Small style to SCV and SCG funds, the Large style to LCV and LCG funds, the Value style to SCV and LCV funds and the Growth style to SCG and LCG funds. We refer to Small, Large, Value and Growth funds as "SLVG" funds for brevity. Fifth, we assign either the Mixed style or the Other style to funds not categorized as corner funds. The Mixed style includes funds that do not meet the threshold for style inclusion defined previously or are assigned to the Lipper codes associated with the Medium row or Blend column of the style box in figure 1.[7] The Other style includes funds with a style that does not fit into the style box.[8]

---

[7] Specifically, these Lipper codes are EIEI, LCCE, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE and SCCE.

[8] The Other funds have Lipper codes EMN, MATC, MATD and MATH.

The previous five steps result in a total of ten style categories: four corner styles (LCG, SCG, LCV and SCV), four SLVG styles (Small, Large, Value and Growth), the Mixed style and the Other style. Panel B of table 4.2 reports the number of funds classified into each style for three values of the threshold for style inclusion: 25%, 50% and 75%.[9] Intuitively, the threshold determines the "style purity" of the corner funds. When its value is 75%, the corner fund numbers are small and vary from 8 small-cap value funds to 42 small-cap growth funds. Most funds end up in the Mixed style because they do not have the required Lipper codes for 75% of their observations. When the threshold is 50%, the corner fund numbers increase considerably and include 53 small-cap value funds, 70 large-cap value funds, 135 small-cap growth funds and 168 large-cap growth funds. The SLVG funds are divided into 188 small-cap funds, 238 large-cap funds, 123 value funds and 303 growth funds. When the threshold is 25%, the fund numbers are above 100 for all corner styles, above 290 for all SLVG styles and equal to 1532 for the Mixed style. There are constantly exactly 205 funds throughout the sample in the Other style.

4.5.3 Style Portfolios and Upper Performance Bound Choice

This section examines the results for portfolios of funds classified into the ten previously identified style categories. The style portfolios constructed from the corner and SLVG funds are particularly important because they are used to extract the style clientele SDFs with the approach discussed in section 4.2.2. To make these portfolios as representative of their style as possible, we select a threshold for style inclusion of 50% in the fund classification method. This threshold ensures that the funds included in the portfolios are assigned to the correct style for a majority of their monthly observations and are in sufficient numbers to provide adequate portfolio diversification (i.e., low fund-specific risk).

Table 4.3 presents summary statistics, including Fama-French SDF alphas and their *t*-statistics, for the monthly returns on two types of portfolios: net asset value-weighted

---

[9] When the threshold is 25%, it is possible for a fund to end up in more than one corner style. To eliminate this possibility, we assign funds to the LCG, SCG, LCV and SCV styles in this predetermined order and remove them from consideration once assigned. As a robustness check, we verify the importance of the ordering for the style classification and find it has little effect as only one fund switches style.

(NAV-weighted) portfolios (panel A) and equally-weighted (EW) portfolios (panel B). The NAV-weighted portfolios in panel A have mean returns from 0.413% to 0.821% and standard deviations from 4.576% to 7.150%. Their Sharpe ratios and Fama-French alphas vary from 0.044 to 0.110 and from -0.084% to 0.193%, respectively. The equally-weighted portfolios in panel B have mean returns from 0.418% to 0.734% and standard deviations from 4.315% to 7.018%. Their Sharpe ratios and Fama-French alphas vary from 0.046 to 0.105 and from -0.098% to 0.146%, respectively. No Fama-French alpha is statistically significant at the 5% level. Overall, the portfolios of small-cap funds, value funds and other funds tend to earn more.

Two findings implied by table 4.3 are useful for the rest of our analysis. First, the results in panels A and B are relatively similar and suggest that the type of weights for the style portfolios is not a material choice. Based on this assessment, we focus on NAV-weighted portfolios, which better account for the clientele invested amounts, to identify the SDFs used in our clientele-specific performance evaluation. In unreported results, we confirm that the SDFs identified from equally-weighted portfolios are similar to those identified from NAV-weighted portfolios.

Second, as discussed in section 4.3.2, the Fama-French alphas are useful to select an appropriate value for the upper performance bound, $\bar{\alpha}_s$, needed to estimate the disagreement parameter of the SDFs identified from corner and SLVG style portfolios. The results in table 4.3 combined with the mean (0.248%) and median (0.212%) bound values reported by Ferson and Lin (2014) allow us to make a selection. Specifically, we find that the clienteles most favorable to the corner and SLVG NAV-weighted style portfolios should expect mean alphas from 0.164% to 0.299% and median alphas from 0.128% to 0.263%. Based on these ranges, we opt for $\bar{\alpha}_s = 0.15\%$ as a realistic yet conservative value for the upper performance bound of the style portfolios. In section 4.6.4, we examine the sensitivity of our results to this choice by considering values of $\bar{\alpha}_s = 0\%$ and $\bar{\alpha}_s = 0.3\%$. These values respectively assume that the style clienteles expect either no performance or a high performance from their preferred style portfolios.

4.6 Empirical Results

4.6.1 Style Clientele Stochastic Discount Factors

We begin our discussion of the empirical results by examining the empirical style clientele SDFs identified from the NAV-weighted style portfolios and an upper performance bound of $\bar{\alpha}_s = 0.15\%$. As discussed in section 4.2.2, the style clientele SDF $\bar{m}_s$ should represent marginal preferences of clienteles favorable to a style. Our style classification method results in eight styles associated with clienteles. The four SLVG styles (Small, Large, Value, Growth) are general and should attract broad clienteles who consider either the size or value focus of equity funds. Accordingly, we refer to the SDFs identified from SLVG style portfolios as one-style clientele (1-SC) SDFs. The four corner styles (LCG, SCG, LCV, SCV) are more specific and should attract specialized clienteles who *jointly* consider the size and value focuses of equity funds. Accordingly, we refer to the SDFs extracted from corner style portfolios as two-style clientele (2-SC) SDFs.

Table 4.4 reports various statistics on the style clientele SDFs. Panel A presents the mean, standard deviation (StdDev) and selected percentiles of the SDFs. As expected, given that a risk-free asset is included in the passive portfolios, all SDFs have the same mean. The standard deviations are from 0.269% to 0.531% for the one-style clientele SDFs and from 0.265% to 0.492% for the two-style clientele SDFs. The lowest values are for the Small and SCV clienteles, and the highest values are for the Large and LCV clienteles. The SDF standard deviations for value versus growth clienteles are closer.

Panel B looks at the correlations between the style clientele SDFs. The results suggest that the broad clienteles identified from the SLVG style portfolios and the specialized clienteles identified from the corner style portfolios are generally different. For the one-style clientele SDFs, the correlations are particularly low between the Large and Value clienteles (Corr. = 0.161), the Value and Growth clienteles (Corr. = 0.395) and the Small and Large clienteles (Corr. = 0.510). On the other hand, the Large and Growth clienteles appear the most related (Corr. = 0.869). For the two-style clientele SDFs, low correlations are found between the LCG and LCV clienteles (Corr. = 0.012), the SCG and LCV clienteles (Corr. = 0.395) and the LCV and SCV clienteles (Corr. = 0.534).

158

The only case where the style portfolios appear to fail to identify economically different clienteles concerns the SCG and SCV SDFs (Corr. = 0.967). Furthermore, these SDFs are almost perfectly correlated with the Small SDFs, suggesting that it is difficult to distinguish between favorable preferences for value versus growth among small-cap fund investors. In fact, given the low volatility and similar statistics of these three SDFs in panel A, they appear closely related to m_{LOP}, the common part to all style clienteles SDFs, suggesting that there is little disagreement between small-cap mutual fund investors.

As discussed by Hansen and Jagannathan (1991), in rational asset pricing theory, an increase in investor risk aversion typically leads to a more volatile SDF. Our results thus suggest that the clienteles with favorable preferences for large-cap funds tend to be more risk averse than the ones attracted to small-cap funds, and that there is little risk aversion difference between the value and growth clienteles. However, in behavioral asset pricing theory, a more volatile SDF can also be the result of behavioral preferences that account for psychological biases and sentiment (Shefrin, 2008, 2009). The higher variability and negative minimum values (which do not preclude arbitrage) for the Large, Growth, LCV and LCG style clientele SDFs are consistent with the evidence of Shefrin and Statman (1995, 2003) and Shefrin (2015) that style investors misjudge the risk-return tradeoff of size and value investments by expecting too high returns (being optimistic) for large-cap and growth stocks and expecting too low returns (being pessimistic) for small-cap and value stocks.

To better understand the rational and behavioral economic properties of the style clientele SDFs, next we empirically project the SDFs on the (net) market portfolio return (r_{mt}) and its squared and cubic values:

(20) $$\bar{m}_{st} = a_0 + a_1 r_{mt} + a_2 r_{mt}^2 + a_3 r_{mt}^3 + e_t,$$

This regression is inspired by the analyses of Harvey and Siddique (2000), Dittmar (2002) and Guidolin and Timmermann (2008), who examine a decomposition of the SDF into a polynomial function of the market return. In particular, Dittmar (2002) starts with the following Taylor series expansion of the SDF:

(21)
$$m = k_0 + k_1 \frac{U''}{U'} R_W + k_2 \frac{U'''}{U'} R_W^2 + \cdots,$$

where U is the investor's utility function and R_W is the return on aggregate wealth. Then he uses the preference theory analyses of Arditti (1967) and Kimball (1993) to sign the first three coefficients of the expansion. Specifically, he shows that positive marginal utility, risk aversion, decreasing absolute risk aversion and decreasing absolute prudence imply that $a_1 < 0$, $a_2 > 0$ and $a_3 < 0$ in rational asset pricing theory. These signs also indicate preferences against risk, for skewness and against kurtosis, respectively.[10]

Although Dittmar (2002) documents evidence in favor of his cubic SDF, he also finds it is not monotone decreasing with the market return, inconsistent with an implication of decreasing absolute risk aversion. In different contexts, Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002) also obtain SDFs that are not monotonically decreasing. Motivated by these findings, Shefrin (2008, 2009) proposes an alternative interpretation of equation (20) from a behavioral asset pricing perspective. He argues that the SDF can be seen as the sum of a component relating to fundamentals (i.e., the rational SDF) and a component relating to sentiment (i.e., the behavioral SDF). In equation (20), the rational component can be thought as the CAPM SDF, $m_{CAPMt} = a_0 + a_1 r_{mt}$, and the behavioral component captures the rest, $\overline{m}_{st} - m_{CAPMt}$. This interpretation provides a role for behavioral preferences in the SDFs of fund investors, consistent with the results of Lee, Shleifer and Thaler (1991), Neal and Wheatley (1998), Goetzmann and Massa (2003), Indro (2004), Bailey, Kumar and Ng (2011) and Blackburn, Goetzmann and Ukhov (2013) that behavioral biases and sentiment are factors in the fund industry.

---

[10] Other rational approaches based on nonlinear decompositions exist to study the economic properties of equity investor SDFs with unknown utility functions. For example, Bansal and Viswanathan (1993) advocate a nonlinear function of the market return and interest rate for the SDF. Bansal, Hsieh and Viswanathan (1993) let the SDF depends on the market return, term spread and interest rate. Chapman (1997) uses functions of aggregate consumption and its lag as state variables in his nonlinear SDF. Rosenberg and Engle (2002) propose an empirical SDF based on a polynomial expansion around the market return that allows time variation in risk aversion. Chabi-Yo (2008) argues for a SDF as a function of payoffs, squared payoffs, skewness and kurtosis of the asset returns.

Figure 2 illustrates the empirical results of the projection in equation (20) for the one-style clientele SDFs (figure 2a) and two-style clientele SDFs (figure 2b). Consistent with a behavioral component, the figure shows that the SDFs are not monotonically decreasing, except perhaps for the Value and LCV clienteles. The SDFs exhibit different preferences, especially when market returns are extreme. When the monthly returns are between -10% and 6% (approximately one to two standard deviations from their mean), the style clientele SDFs are generally similar and show a negative slope consistent with rational asset pricing theory. This segment captures the rational SDF component and suggests that clienteles have similar risk aversion in "normal" market states.

However, when the market states are extreme, the SDFs significantly deviate from each other and the results are more consistent with a role for irrationality and market sentiment (e.g., Lakonishok, Shleifer and Vishny, 1994, Baker and Wurgler, 2006, 2007). In particular, figure 1a shows that the Growth and Large SDFs are *increasing* with returns in extreme market states. In difficult (easy) times, when market sentiment is low (high), the Growth and Large clienteles have relatively low (high) SDFs and are thus less (more) attracted that they rationally should be to their style funds, an interpretation also valid for the LCG clienteles in figure 1b. In contrast, the Value and (to a lesser extent) Small clienteles have relatively high (low) SDFs in difficult (easy) times, when market sentiment is low (high), and are thus more (less) attracted that they rationally should be to their style funds. The same interpretation holds for the LCV clienteles.

Overall, figure 2 indicates that the Growth and Large SDFs include a significant behavioral component because they appear to underweight the probabilities of extreme negative events and overweight the probabilities of extreme positive events, leading them to follow market sentiment. Hence, the clienteles of growth and large-cap funds tend to be optimists and trend followers. The Value and Small SDFs conform more closely to rational asset pricing theory, but their associated clienteles still show a tendency for pessimism and being contrarian, especially for value fund investors. Our style clientele SDFs thus have features that are consistent with the results of Shefrin and Statman (1995, 2003), Blackburn, Goetzmann and Ukhov (2013) and Shefrin (2015), who study the judgments, sentiment sensitivity and trading behavior of style investors. Figure 2 also supports the

findings from table 4.4 that the identified clienteles are generally different (except for the SCG and SCV clienteles) and thus should generate meaningful disagreement in performance evaluation.

4.6.2   Style-Clientele-Specific Performance Evaluation Results

This section examines clientele-specific performance evaluation using the style clientele SDFs. As discussed in section 4.2.3, the evaluation necessitates that we distinguish between individual mutual funds with and without styles associated to clienteles. In our classification, the funds assigned to styles associated with clienteles include the SLVG funds and the corner funds. Their performance measurement simply uses their associated style clientele SDFs, i.e., the one-style clientele SDFs for SLVG funds and the two-style clientele SDFs for corner funds. The funds classified in the Mixed and Other categories represent the funds not assigned to styles associated with clienteles. As they cannot be assigned any unique style clientele SDF, their performance measurement is done for all clienteles.

To obtain a sufficiently large cross-section of funds in each style category, we select a threshold for style inclusion of 25% in the fund classification method. Intuitively, we thus stipulate that a fund assigned to a given style for at least 25% of its return observations should be of interest to the clienteles represented by its associated style clientele SDF. Given this threshold, the number of funds in each cross-section is given in the row under 25% in panel B of table 4.2 and vary from 102 SCV funds to 1532 Mixed funds.

Tables 4.5 to 4.10 present the style-clientele specific performance evaluation results. Panel A of each table provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated SDF alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the cross-sectional mean of estimated alphas using the test described in section 4.3.3, which accounts for the dependence in performance between funds. Panel B of each table gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$).

162

Table 4.5 examines the results for the broader clienteles who consider either the size or value focus of equity funds. On average, small-cap, large-cap and value funds provide a relatively neutral performance to their respective clienteles, while growth funds provide a significantly positive performance to growth clienteles. The mean alphas are equal to 0.000% (*t*-stat. = 0.00) for small-cap funds, 0.004% (*t*-stat. = 0.06) for large-cap funds, 0.058% (*t*-stat. = 1.00) for value funds and 0.212% (*t*-stat. = 2.43) for growth funds. The percentile statistics confirm that the alpha distributions are centered at approximately zero for small-cap, large-cap and value funds, although the *t*-statistic distributions show more significantly positive than negative alphas for large-cap and value funds. For growth funds, the alpha distribution is centered at a positive value and the proportions in panel B show that the fraction of positive alphas (65.34%) is almost twice the fraction of negative alphas (34.66%).

Table 4.6 provide the results for the more specialized clienteles who jointly consider the size and value focuses of funds. The alphas have means of 0.076% (*t*-stat. = 1.31) for large-cap growth funds, 0.066% (*t*-stat. = 0.97) for small-cap growth funds, 0.101% (*t*-stat. = 2.24) for large-cap value funds and 0.038% (*t*-stat. = 0.93) for small-cap value funds. Although the mean alphas are not significant for the funds assigned to three of the four styles, the percentiles of the distributions in panel A and the proportions in panel B indicate more positive than negative alphas for all four styles. Overall, tables 4.5 and 4.6 suggest that funds assigned to styles associated with clienteles do not underperform for their clienteles. The performance of mutual funds could be more positive than existing evidence shows if the evaluation considers the relevant clienteles.

Tables 4.7 and 4.8 give the results for the 1532 Mixed mutual funds using the one-style clientele SDFs and the two-style clientele SDFs, respectively. These funds tend to fall in the Medium or Blend category of the mutual fund style box and so could presumably be partly attractive for the style clienteles we investigate. In table 4.7, the mean alphas vary from -0.149% (*t*-stat. = -1.98) to 0.020% (*t*-stat. = 0.24) and the proportions of positive alphas vary from 28.20% to 45.89%. In table 4.8, the mean alphas vary from -0.312% (*t*-stat. = -3.02) to -0.056% (*t*-stat. = -0.77) and the proportions of positive alphas vary from 23.24% to 36.75%. The performance of Mixed funds thus tends to be negative.

Tables 4.9 and 4.10 give the results for the 205 Other mutual funds using the one-style clientele SDFs and the two-style clientele SDFs, respectively. These funds are difficult to categorize because they do not fit into the mutual fund style box. In table 4.9, the mean alphas vary from 0.022% (*t*-stat. = 0.37) to 0.200% (*t*-stat. = 3.09) and the proportions of positive alphas vary from 45.85% to 76.59%. In table 4.10, the mean alphas vary from -0.067% (*t*-stat. = -0.90) to 0.119% (*t*-stat. = 2.00) and the proportions of positive alphas vary from 44.39% to 63.90%. The performance of Other funds is thus generally more positive.

Overall, tables 4.7 to 4.10 find that funds not assigned to styles associated with clienteles have performance sensitive to the SDFs used for evaluation. The mean performance values for Large, Growth and LCG clienteles are neutral for Mixed funds and significantly positive for Other funds. The mean values for Small, Value and LCV clienteles are significantly negative for Mixed funds and neutral for Other funds. As discussed in the previous section, the clienteles of large-cap and growth funds tend to be optimistic trend followers and the clienteles of small-cap and value funds tend to be pessimistic contrarians. Our results for Mixed and Other funds thus suggest that the behavioral features of the investor SDFs are important determinants of clientele-specific performance evaluation.

4.6.3   Value Added with Style-Clientele-Specific Performance Measures

This section studies the value added of the actively-managed fund industry from the perspective of the style clienteles by combining the categorized funds into a full-sample cross-section. In the style-clientele-specific evaluation of the previous section, each individual mutual fund is given multiple performance values. Each corner fund is evaluated with three SDFs, i.e., two one-style clientele SDFs and one two-style clientele SDF. For example, a large-cap growth fund is evaluated with the Large, Growth and LCG SDFs. Each Mixed or Other fund is evaluated with the eight style clientele SDFs. To obtain a full-sample cross-sectional distribution of the alphas where each fund is only included once, we thus need to select one alpha per fund.

164

Instead of subjectively picking one value, this section examines many cross-sectional distributions by considering either the minimum or maximum alpha for each fund, thus providing a range of performance cross-sections accounting for various clienteles. Given the previous findings that alphas tend to vary with the behavioral features of the style clientele SDFs, the minimum (maximum) alpha distribution can be roughly interpreted as the mutual fund industry performance for investors who tend to be pessimistic contrarians (optimistic trend followers).

Table 4.11 presents the results for the minimum (Min) and maximum (Max) alpha distributions, with funds evaluated either with the one-style clientele SDFs (under 1-SC alphas) or the two-style clientele SDFs (under 2-SC alphas). Figure 3 illustrates these distributions. As in tables 4.5 to 4.10, panel A provides statistics on the distributions of the estimated SDF alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics), including the t-statistics (t-stat) on the significance of the cross-sectional mean of estimated alphas. In panel B, instead of providing proportions of positive and negative alphas as in previous tables, we report the proportions of estimated alphas that are larger ($\%\alpha_{MF,S} > \alpha_{FF}$) and smaller ($\%\alpha_{MF,S} < \alpha_{FF}$) than the Fama-French estimated alphas (given in table 4.1), which allow us to compare the value added for style clienteles to the value added from a widely-used model.

The results show that the sign of the value added by the fund industry is ambiguous and depends on the choice of measures. The minimum one-style and two-style clientele alpha estimates show a negative average performance of -0.182% (t-stat. = -2.71) and -0.239% (t-stat. = -2.68), respectively, and are below the Fama-French alphas for more than 70% of funds. In contrast, the maximum one-style and two-style clientele alpha estimates show a positive average performance of 0.174% (t-stat. = 2.29) and 0.074% (t-stat. = 1.21), respectively, and are above the Fama-French alphas for more than 75% of funds. The disagreement in alpha is well illustrated by the alpha distributions in figure 3. Overall, these findings suggest that the value added by the fund industry tend to be negative for investors with pessimistic-contrarian behavioral SDFs, but positive for those with optimistic-trend-following behavioral SDFs.

4.6.4 Sensitivity of Value Added Results to Upper Performance Bound Choice

Our empirical results have thus far relied on the style clientele SDFs identified with an upper performance bound for the style portfolios of $\bar{\alpha}_s = 0.15\%$. This section examines the sensitivity of our results to this choice by considering values of $\bar{\alpha}_s = 0\%$ and $\bar{\alpha}_s = 0.3\%$. Compared to the value of $\bar{\alpha}_s = 0.15\%$ which is consistent with the findings of Ferson and Lin (2014), these values respectively assume that the style clienteles expect either no performance or a relatively high performance from their preferred style portfolios.

Tables 4.12 and 4.13 reproduce the value added results of table 4.11 by using $\bar{\alpha}_s = 0\%$ and $\bar{\alpha}_s = 0.3\%$, respectively. The tables show that the investor disagreement observed from the difference between the minimum and maximum alphas increases with the upper performance bound. Furthermore, this increase is caused by movements in the maximum alpha distributions. Specifically, the results for the minimum alpha distributions are similar to those in table 4.11, including comparable negative mean alphas and high proportions of clientele alphas below the Fama-French alphas. In contrast, the results for the maximum alpha distributions are relatively neutral when $\bar{\alpha}_s = 0\%$, but more positive when $\bar{\alpha}_s = 0.3\%$ than when $\bar{\alpha}_s = 0.15\%$. When $\bar{\alpha}_s = 0\%$, the maximum alpha estimates have insignificant means and are above the Fama-French alphas for only an average of 55% of funds. When $\bar{\alpha}_s = 0.3\%$, the maximum alpha estimates have highly statistically significant means and are above the Fama-French alphas for 85% of funds. Overall, these results confirm that the sign of the value added by the fund industry is generally ambiguous and depends on the choice of clienteles. But they also show that it partly depends on the maximum performance expected by the clienteles most favorable to the style.

4.7 Conclusion

Mutual funds cater to and attract specific clienteles throughout their investment style. We propose clientele-specific performance measures based on the implied style preferences of mutual fund investors. The performance framework is based on a SDF alpha approach with investor disagreement, following Chrétien and Kammoun (2015). The identification of meaningful SDFs for style clienteles uses representative style portfolios and the findings of Ferson and Lin (2014) on investor disagreement. The style classification employs a new

method to better exploit existing objective code data from Lipper and to account for code changes and missing codes.

Our empirical investigation uses a sample of 2530 U.S. equity mutual funds with monthly returns from 1998 to 2012. The economic properties of the SDFs for style clienteles indicate that the preferences implied by the SDFs have similar risk aversion but differ in their behavioral features. Value and small-cap SDFs show pessimism and contrarian behavior. Growth and large-cap SDFs show optimism and trend following behavior. The style-clientele-specific performance evaluation finds that funds assigned to equity styles have a neutral to positive performance when they are evaluated with their relevant clientele-specific measure. The performance of the other funds is sensitive to the clienteles and the behavioral features of the SDFs are important determinants of the evaluation. The value added by the mutual fund industry also depends on the choice of measures. Overall, we find that preferences and performance evaluations differ for size and value mutual fund clienteles. We provide supporting evidence for the conjecture of Ferson (2010) that clientele-specific measures based on meaningful investor clienteles might be necessary to properly evaluate mutual funds.

We agree with Ferson (2010) and Ferson and Lin (2014) on their calls for more research on clientele effects in performance evaluation. Our clientele-specific performance approach can serve as a useful framework for developing measures to account for other clientele effects documented by the literature. Future research can also employ different techniques to more fully characterize the economic properties of preferences implied by the clientele SDFs. Finally, other strategies based on different restrictions or alternative incomplete market setups can be followed to identify economically meaningful clientele SDFs useful for clientele-specific measurement.

4.8 References

Ahn, D., Cao, H. H., Chrétien, S., (2009), "Portfolio performance measurement: a no arbitrage bounds approach", *European Financial Management*, 15, 2, 298-339.

Aït-Sahalia, Y., Lo, A. W., (2000), "Nonparametric risk management and implied risk aversion", *Journal of Econometrics*, 94, 9-51.

Arditti, F. D., (1967), "Risk and the required return on equity", *Journal of Finance*, 22, 19-36.

Bailey, W., Kumar, A., Ng, D., (2011), "Behavioral biases and mutual fund investors", *Journal of Financial Economics*, 102, 1-27.

Baker, J., Wurgler, J., (2006), "Investor sentiment and the cross-section of stock returns", *Journal of Finance*, 61, 1645-1680.

Baker, J., Wurgler, J., (2007), "Investor sentiment in the stock market ", *Journal of Economic Perspectives*, 21, 129-151.

Bansal, R., Hsieh, D. A., Viswanathan, S., (1993), "A new approach to international arbitrage pricing", *Journal of Finance,* 48, 5, 1719-1747.

Bansal, R., Viswanathan, S., (1993), "No-arbitrage and arbitrage pricing: A new approach", *Journal of Finance,* 48, 1231-1262.

Barber, B. M., Odean, T., Zheng, L., (2005), "Out of sight, out of mind: the effects of expenses on mutual fund flows", *Journal of Business*, 78, 2095-2120.

Barras, L., Scaillet, O., Wermers, R., (2010), "False discoveries in mutual fund performance: Measuring luck in estimated alphas", *Journal of Finance*, 65, 179-216.

Barberis, N., Shleifer, A., (2003), "Style investing", *Journal of Financial Economics*, 68, 161-199.

Bergstresser, D., Chalmers, J., Tufano, P., (2009), "Assessing the costs and benefits of brokers in the mutual fund industry", *Review of Financial Studies*, 22, 4129-4156.

Blackburn, D. W., Goetzmann, W. N., Ukhov, A. D., (2013), "Is trading behavior stable across contexts? Evidence from style and multi-style investors", *Quantitative Finance*, 1-23.

Brown, S. J., Goetzmann, W. N., (1997), "Mutual fund styles", *Journal of Financial Economics*, 43, 373-399.

Chrétien, S., Kammoun, M., (2015), "Mutual fund performance evaluation and best clientele", Unpublished Working Paper, Laval University.

Chabi-Yo, F., (2008), "Conditioning information and variance bounds on pricing kernels with higher-order moments: theory and evidence", *Review of Financial Studies*, 21, 1, 181-231.

Chapman, D. A., (1997), "Approximating the asset pricing kernel", *Journal of Finance*, 52, 4, 1383-1410.

Chen, Z., Knez, P. J., (1996), "Portfolio performance measurement: theory and applications", *Review of Financial Studies*, 9, 2, 511-555.

Christoffersen, S., Evans, R. Musto, D., (2013), "What do consumers' fund flows maximize? Evidence from their brokers' incentives", *Journal of Finance*, 68, , 201-235.

Cochrane, J., Saá-Requejo, J., (2000), "Beyond arbitrage: good-deal asset price bounds in incomplete markets", *Journal of Political Economy,* 108, 79-119.

Dahlquist, M., Söderlind, P., (1999), "Evaluating portfolio performance with stochastic discount factors", *Journal of Business*, 72, 347-383.

Del Guercio, D., Reuter, J., (2014), "Mutual fund performance and the incentive to generate alpha", *Journal of Finance,* 69, 1673-1704.

Del Guercio, D., Tkac, P. A., (2002), "The determinants of the flow of funds of managed portfolios: mutual funds vs. pension funds", *Journal of Financial and Quantitative Analysis*, 37, 4, 523-557.

Dibartolomeo, D., Witkowski, E., (1997), "Mutual fund misclassification: evidence based on style analysis", *Financial Analysts Journal*, 53, 5, 32-43.

Dittmar, R., (2002), "Nonlinear pricing kernels, kurtosis preferences and the cross-section of equity returns", *Journal of Finance*, 57, 369-403.

Elton, E. J., Gruber, M. J., Blake, C. R., (2001), "A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases", *Journal of Finance*, 56, 2415-2430.

Evans, R.B., (2010), "Mutual fund incubation", *Journal of Finance*, 65, 1581-1611.

Fama, E. F., (1998), "Market efficiency, long-term returns and behavioral finance", *Journal of Financial Economics*, 49, 283-306.

Fama, E. F., French, K. R., (1992), "The cross-section of expected stock returns", *Journal of Finance*, 47, 427-465.

Fama, E. F., French, K. R., (1993), "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, 33, 3-56.

Fama, E. F., French, K. R., (2010), "Luck versus skill in the cross-section of mutual fund returns", *Journal of Finance*, 65, 1915-1948.

Farnsworth, H., Ferson, W., Jackson, D., Todd, S., (2002), "Performance evaluation with stochastic discount factors", *Journal of Business*, 75, 473-503.

Ferson, W, E., Chen, Y., (2015), "How many good and bad fund managers are there, really?", Unpublished Working Paper, University of Southern California and Texas A&M University.

Ferson, W. E., (2010), "Investment performance evaluation", *Annual Review of Financial Economics*, 2, 207-234.

Ferson, W. E., Lin, J., (2014), "Alpha and performance measurement: the effect of investor heterogeneity", *Journal of Finance,* 69, 1565-1596.

Glode, V., (2011), "Why mutual funds underperform", *Journal of Financial Economics*, 93, 546-559.

Glosten, L., Jagannathan, R., (1994), "A contingent claims approach to performance evaluation", *Journal of Empirical Finance,* 1, 133-66.

Goetzmann, W. N., Massa, M., (2002), "Daily momentum and contrarian behavior of index funds investors", *Journal of Financial and Quantitative Analysis*, 37, 375-389.

Goetzmann, W. N., Massa, M., (2003), "Index funds and stock market growth", *Journal of Business*, 76, 1, 1-28.

Guidolin, M., Timmermann, A., (2008), "International asset allocation under regime switching, skew and kurtosis preferences", *Review of Financial Studies*, 21, 889-935.

Hansen, L. P., (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029-1054.

Hansen, L. P., Jagannathan, R., (1991), "Implication of security market data for models of dynamic economies", *Journal of Political Economy*, 99, 225-262.

Harvey, C. R., Siddique, A., (2000), "Conditional skewness in asset-pricing tests", *Journal of Finance*, 55, 1263-1295.

Indro, D. C., (2004), "Does mutual fund flow reflect investor sentiment?", *Journal of Behavioral Finance*, 5, 105-115.

Ivković, Z., Weisbenner, S., (2009), "Individual investors mutual fund flows", *Journal of Financial Economics*, 92, 223-237.

James, C., Karceski, J., (2006), "Investor monitoring and differences in mutual fund performance", *Journal of Banking and Finance*, 30, 2787-2808.

Kacperczyk, M., Sialm, C., Zheng, L., (2008), "Unobserved actions of mutual funds", *Review of Financial Studies*, 21, 2379-2416.

Kim, M., Shukla, R., Tomas, M., (2000), "Mutual fund objective misclassification", *Journal of Economics and Business*, 52, 309-323.

Kimball, M., (1993), "Standard risk aversion", *Econometrica*, 61, 589-611.

Lakonishok, J., Shleifer, A., Vishny, R. W., (1994), "Contrarian investment, extrapolation, and risk", *Journal of Finance*, 49, 1541-1578.

Lee, C. M. C., Shleifer, A., Thaler, R. H., (1991), "Investor sentiment and the closed-end fund puzzle", *Journal of Finance*, 46, 75-109.

Malloy, C., Zhu, N., (2004), "Mutual fund choices and investor demographics", Unpublished Working Paper, University of California at Davis.

Nanda, V., Narayanan, M. P., Warther, V., (2000), "Liquidity, investment ability, and mutual fund structure", *Journal of Financial Economics*, 57, 417-443.

Neal, R., Wheatley, S. M., (1998), "Do measures of investor sentiment predicts returns?", *Journal of Financial and Quantitative Analysis*, 33, 523-547.

Rosenberg, J. V., Engle, R. F., (2002), "Empirical pricing kernel", *Journal of Financial Economics*, 64, 341-372.

Pástor, Ĺ., Stambaugh, R., (2002), "Mutual fund performance and seemingly unrelated assets", *Journal of Financial Economics*, 63, 315-349.

Sharpe, W. F., (1992), "Asset allocation: management style and performance measurement", *Journal of Portfolio Management*, 18, 7-19.

Shefrin, H., (2008), *A Behavioral Approach to Asset Pricing*, 2nd edition, Elsevier Academic Press, NY, NY.

Shefrin, H., (2009), "Behavioralizing finance", *Foundations and Trends in Finance*, 4, 1-2, 1-184.

Shefrin, H., (2015), "Investors' judgments, asset pricing factors and sentiment", *European Financial Management*, 2, 205-227.

Shefrin, H., Statman, M., (1995), "Making sense of beta, size and book-to-market", *Journal of Portfolio Management*, 21, 26-34.

Shefrin, H., Statman, M., (2003), "The style of investor expectations", in Coggin, T., Fabozzi, F. (eds), *Handbook of Equity Style Management*, Wiley, 195-218.

Table 4.1: Summary Statistics

Table 4.1 presents summary statistics for the monthly data from January 1998 to December 2012. Panel A shows cross-sectional summary statistics (average (Mean), standard deviation (StdDev) and selected percentiles) on the distributions of the average (Mean), standard deviation (StdDev), minimum (Min), maximum (Max), Sharpe ratio (h) and Fama-French SDF Alphas with their corresponding t-statistics for the returns on 2530 actively-managed open-ended U.S. equity mutual funds. It also reports the t-statistics (t-stat) on the significance of the mean of estimated Fama-French SDF alphas (see test description in section 4.3.3). Panel B gives the average (Mean), standard deviation (StdDev), minimum (Min), maximum (Max) and Sharpe ratio (h) for the passive portfolio returns. The passive portfolios include ten industry portfolios (consumer nondurables (NoDur), consumer durables (Dur), manufacturing (Manuf), energy (Enrgy), high technology (HiTec), telecommunication (Telcm), shops (Shops), healthcare (Hlth), utilities (Utils), and other industries (Other)), and the risk-free asset (RF) based on the one-month Treasury bills. All statistics are in percentage except for the Sharpe ratios and the t-statistics.

Panel A: Mutual Fund Returns and Fama-French Alphas

	Mutual Fund Returns					Fama-French Alphas	
	Mean	StdDev	Min	Max	h	α_{MF}	t-statistics
Mean	0.4848	5.6633	-19.1187	16.1357	0.0452	-0.0637	-0.3924
StdDev	0.3091	1.6530	4.9232	7.5524	0.0547	0.2607	1.2285
(t-stat)						(-1.161)	
Max	1.8164	17.6776	-2.1401	75.0000	0.2324	1.3163	4.1081
99%	1.1755	10.5577	-5.2848	40.5801	0.1537	0.5577	2.2049
95%	0.9323	8.5988	-12.7480	32.1314	0.1220	0.2798	1.5057
90%	0.8391	7.6060	-14.2134	26.8592	0.1061	0.1988	1.1115
75%	0.6623	6.3709	-16.2791	18.3857	0.0791	0.0698	0.4345
Median	0.4897	5.3792	-18.6207	13.8677	0.0517	-0.0517	-0.3555
25%	0.3410	4.7017	-21.7059	11.3099	0.0207	-0.1723	-1.1245
10%	0.1223	4.1858	-24.6343	9.7340	-0.0264	-0.3199	-1.9346
5%	-0.0178	3.7186	-26.4187	8.8372	-0.0550	-0.4475	-2.5100
1%	-0.3862	1.6135	-35.9893	5.0661	-0.1179	-0.9014	-3.8027
Min	-2.5734	0.9172	-59.0909	2.4218	-0.3063	-3.2906	-6.1673

Panel B: Passive Portfolio Returns

	Mean	StdDev	Min	Max	h
NoDur	0.6734	3.8159	-13.0900	10.9000	0.1226
Durbl	0.5574	8.1845	-32.8900	42.9200	0.0431
Manuf	0.7896	5.4291	-20.9400	17.7800	0.1077
Enrgy	0.9989	6.1105	-17.1200	19.1300	0.1301
HiTec	0.7310	8.2327	-26.1500	20.4600	0.0640
Telcm	0.4103	5.9385	-15.5600	22.1200	0.0346
Shops	0.7485	4.8399	-15.1600	13.3800	0.1122
Hlth	0.6037	4.2342	-12.3600	12.0300	0.0944
Utils	0.6949	4.4358	-12.6500	11.7600	0.1107
Other	0.3556	5.7347	-21.2800	16.1100	0.0264
RF	0.2047	0.1721	0.0000	0.5600	-

Table 4.2: Style Classification of Mutual Funds

Table 4.2 presents summary statistics for the style classification of mutual funds. Panel A is based on the Lipper classification codes in the CRSP mutual fund database. It gives the average (Mean), standard deviation (StdDev) and selected percentiles of the numbers of code changes per fund (Nb Changes), the numbers of different codes per fund (Nb Styles), the maximum proportions of monthly observations that a fund has the same code (% Max) and the proportions of monthly observations that a fund presents a missing code (% Missing). Panel B reports the number of mutual funds classified into each style using the style classification method described in section 4.5.2 and three values of the threshold for style inclusion: 25%, 50% and 75%. The classification depends on whether the proportions of monthly return observations associated with the Lipper codes attached to the funds are above the given threshold for style inclusion. Funds are classified into ten styles: four corner styles (LCG, SCG, LCV and SCV), four SLVG styles (Small, Large, Value and Growth), the Mixed style and the Other style. The corner styles include large-cap growth funds (LCG), small-cap growth funds (SCG), large-cap value funds (LCV), small-cap value funds (SCV). The SLVG styles include small-cap funds (Small), large-cap funds (Large), value funds (Value) and growth funds (Growth). The data cover the period January 1998-December 2012.

Panel A: Lipper Classification Codes				
	Nb Changes	Nb Styles	% Max	% Missing
Mean	3.77	3.19	67.46	12.83
StdDev	2.01	1.00	17.12	6.09
Max	13	11	100.00	47.40
99%	10	6	93.00	31.60
95%	8	5	89.83	25.75
90%	6	4	89.83	22.24
75%	5	4	82.30	14.64
Median	3	3	69.45	10.18
25%	2	3	53.61	10.18
10%	2	2	43.84	7.72
5%	1	2	38.98	7.00
1%	1	2	30.51	0.00
Min	0	1	19.08	0.00

Panel B: Style Classification of Mutual Funds			
Style Classification	25%	50%	75%
LCG	298	168	41
SCG	204	135	42
LCV	189	70	14
SCV	102	53	8
Small	306	188	50
Large	487	238	55
Value	291	123	22
Growth	502	303	83
Mixed	1532	1899	2220
Other	205	205	205

Table 4.3: Summary Statistics for Mutual fund Style Portfolios

Table 4.3 presents summary statistics for the monthly returns on net asset value-weighted (NAV-weighted) (panel A) and equally-weighted (EW) (panel B) style portfolios. The table shows the average (Mean), standard deviation (StdDev), minimum (Min), maximum (Max), Sharpe ratio (h) and Fama-French SDF alphas with their corresponding t-statistics of the portfolio returns. Using the style classification method described in section 4.5.2 and a 50% threshold for style inclusion, the funds are classified into ten styles: four corner styles (LCG, SCG, LCV and SCV), four SLVG styles (Small, Large, Value and Growth), the Mixed style and the Other style. The corner styles include large-cap growth funds (LCG), small-cap growth funds (SCG), large-cap value funds (LCV), small-cap value funds (SCV). The SLVG styles include small-cap funds (Small), large-cap funds (Large), value funds (Value) and growth funds (Growth). One NAV-weighted portfolio and one EW portfolio are then formed for each style. The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except for the Sharpe ratios and the t-statistics.

Panel A : Returns and Fama-French Alphas for NAV-Weighted Portfolios							
	NAV-Weighted					Fama-French Alphas	
Style Classification	Mean	StdDev	Min	Max	h	α_S	t-statistics
LCG	0.4482	5.4292	-16.8991	12.4339	0.0477	-0.0720	-0.9200
SCG	0.7082	7.1505	-22.0506	27.0381	0.0707	-0.0830	-0.7800
LCV	0.4128	4.6562	-16.9076	11.7733	0.0450	-0.0720	-0.9200
SCV	0.8206	5.5741	-21.3388	17.7120	0.1084	0.0514	0.4200
Small	0.7691	6.5758	-21.0640	23.3101	0.0851	-0.0350	-0.4000
Large	0.4169	5.0808	-16.8712	11.7631	0.0437	-0.0840	-1.4700
Value	0.5760	4.9469	-19.1288	14.1630	0.0745	-0.0080	-0.0900
Growth	0.4933	5.6334	-17.9827	13.6074	0.0534	-0.0610	-0.8200
Mixed	0.5450	4.8643	-18.0761	11.3832	0.0692	-0.0030	-0.0900
Other	0.7079	4.5765	-18.6406	10.6618	0.1103	0.1927	1.9200
Panel B : Returns and Fama-French Alphas for Equally-Weighted Portfolios							
	EW					Fama-French Alphas	
Style Classification	Mean	StdDev	Min	Max	h	α_S	t-statistics
LCG	0.4412	5.3274	-16.9975	12.2806	0.0457	-0.0680	-0.8100
SCG	0.6979	7.0176	-21.4409	24.1058	0.0690	-0.0980	-0.9800
LCV	0.4179	4.5526	-16.4539	11.6782	0.0461	-0.0680	-0.8100
SCV	0.7344	5.3197	-20.6288	16.1120	0.0975	0.0122	0.0900
Small	0.7078	6.3206	-20.9842	17.8977	0.0781	-0.0680	-0.7800
Large	0.4292	4.9289	-16.8034	11.6983	0.0464	-0.0710	-1.4500
Value	0.5517	4.7368	-18.1392	12.8513	0.0718	-0.0340	-0.3400
Growth	0.5566	5.9526	-18.9806	15.2024	0.0591	-0.0810	-1.1300
Mixed	0.5165	4.9201	-18.0184	11.9546	0.0627	-0.0580	-1.1300
Other	0.6548	4.3148	-18.9360	9.9330	0.1048	0.1463	1.4300

174

Table 4.4: Statistics for the Style Clientele Stochastic Discount Factors

Table 4.4 shows statistics for one-style and two-style clientele stochastic discount factors (denoted 1-SC SDFs and 2-SC SDFs, respectively). The one-style clientele SDFs are estimated from NAV-weighted style portfolios of small-cap funds (Small), large-cap funds (Large), value funds (Value) and growth funds (Growth). The two-style clientele SDFs are estimated from NAV-weighted style portfolios of large-cap growth funds (LCG), small-cap growth funds (SCG), large-cap value funds (LCV) and small-cap value funds (SCV). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the SDFs. Panel B provides correlations between the SDFs. The portfolio data (see statistics in table 4.3) cover the period January 1998-December 2012.

Panel A : Distributions of SDFs Extracted from NAV-Weighted Portfolios								
	1-SC SDFs				2-SC SDFs			
	Small	Large	Value	Growth	LCG	SCG	LCV	SCV
Mean	0.9979	0.9980	0.9980	0.9980	0.9980	0.9979	0.9979	0.9979
StdDev	0.2691	0.5307	0.3456	0.3926	0.4202	0.2768	0.4920	0.2654
Max	1.6976	2.5202	2.0162	2.2619	2.2416	1.7253	2.4430	1.7489
99%	1.6797	2.2689	1.8602	2.1642	2.2094	1.6767	2.4272	1.6545
95%	1.4818	1.9003	1.5829	1.5972	1.6521	1.5001	1.7381	1.4777
90%	1.3841	1.6542	1.4490	1.4444	1.4883	1.3887	1.6597	1.3803
75%	1.1552	1.3703	1.2136	1.2453	1.2848	1.1484	1.2686	1.1564
Median	0.9778	0.9760	1.0111	0.9949	0.9751	0.9744	0.9639	0.9713
25%	0.8151	0.6718	0.7747	0.7747	0.7569	0.8135	0.6924	0.8199
10%	0.6719	0.3740	0.5420	0.5290	0.5354	0.6497	0.4243	0.6822
5%	0.5968	0.0843	0.4130	0.3924	0.2996	0.5750	0.2566	0.6108
1%	0.4008	-0.3958	0.2733	-0.0868	-0.0936	0.3475	-0.2323	0.3967
Min	0.3515	-0.8958	0.2459	-0.3450	-0.4548	0.3343	-0.6500	0.3881

Panel B: Correlations between SDFs Extracted from NAV-Weighted Portfolios								
	Small	Large	Value	Growth	LCG	SCG	LCV	SCV
Small	1.0000	0.5096	0.7929	0.7389	0.6375	0.9919	0.4654	0.9910
Large		1.0000	0.1606	0.8691	0.9575	0.5361	0.0351	0.4849
Value			1.0000	0.3949	0.2323	0.7726	0.6517	0.7943
Growth				1.0000	0.9266	0.7920	-0.0214	0.6796
LCG					1.0000	0.6597	0.0119	0.6160
SCG						1.0000	0.3951	0.9673
LCV							1.0000	0.5336
SCV								1.0000

Table 4.5: One-Style Clientele Alphas for SLVG Mutual Funds

Table 4.5 shows statistics on the cross-sectional distribution of monthly SDF alphas for SLVG funds using the one-style clientele SDFs. The one-style clientele SDFs are estimated from NAV-weighted style portfolios of small-cap funds (Small), large-cap funds (Large), value funds (Value) and growth funds (Growth). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$). The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

	Panel A: Performance and t-statistics							
	Performance				t-statistics			
	Small	Large	Value	Growth	Small	Large	Value	Growth
Mean	0.0003	0.0044	0.0579	0.2123	-0.0176	0.1250	0.0777	0.2019
StdDev	0.2811	0.3715	0.2672	0.4083	0.7277	0.3821	0.4098	0.4405
(t-stat)	(0.005)	(0.056)	(0.997)	(2.426)				
Max	0.7067	1.5898	0.9179	1.4943	2.3518	2.6492	3.0433	2.3506
99%	0.5759	1.0822	0.6942	1.1197	2.1082	1.3675	2.5134	1.7246
95%	0.4225	0.6395	0.5217	0.9628	1.0840	0.7924	0.6433	0.9862
90%	0.3448	0.4726	0.4330	0.7848	0.7587	0.5473	0.3866	0.6250
75%	0.1711	0.2392	0.2661	0.4934	0.2470	0.3511	0.2202	0.4109
Median	0.0098	-0.0219	-0.0164	0.1424	0.0122	-0.0221	-0.0106	0.2102
25%	-0.1330	-0.2489	-0.1382	-0.0756	-0.2690	-0.1245	-0.1145	-0.0950
10%	-0.3498	-0.4587	-0.2158	-0.2661	-0.7908	-0.1951	-0.2198	-0.2545
5%	-0.4884	-0.5495	-0.2962	-0.3806	-1.1516	-0.2464	-0.3326	-0.3920
1%	-0.8076	-0.7302	-0.4452	-0.5979	-2.3802	-0.5447	-0.7393	-0.7063
Min	-1.2032	-0.9009	-0.6669	-0.9238	-2.7732	-1.3152	-0.9996	-1.2347

Panel B: Performance Proportions					
		Small	Large	Value	Growth
Performance	$\%\alpha_{MF,s} > 0$	51.63	47.84	47.42	65.34
Sign	$\%\alpha_{MF,s} < 0$	48.37	52.16	52.58	34.66

Table 4.6: Two-Style Clientele Alphas for Corner Mutual Funds

Table 4.6 shows statistics on the cross-sectional distribution of monthly SDF alphas for corner funds using the two-style clientele SDFs. The two-style clientele SDFs are estimated from NAV-weighted style portfolios of large-cap growth funds (LCG), small-cap growth funds (SCG), large-cap value funds (LCV) and small-cap value funds (SCV). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$). The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

	Panel A: Performance and t-statistics							
	Performance				t-statistics			
	LCG	SCG	LCV	SCV	LCG	SCG	LCV	SCV
Mean	0.0758	0.0663	0.1005	0.0378	0.1922	0.2607	0.2620	0.0769
StdDev	0.2670	0.3101	0.2022	0.1758	0.5610	0.8876	0.4938	0.6741
(t-stat)	(1.3069)	(0.9690)	(2.2447)	(0.9320)				
Max	1.1392	0.8749	0.6930	0.4862	2.2038	2.6301	2.4769	1.6074
99%	0.7725	0.6959	0.6600	0.4333	2.0107	2.5024	2.0237	1.5501
95%	0.5591	0.6017	0.4800	0.3120	1.1447	1.8948	1.4005	1.0487
90%	0.4089	0.4993	0.4156	0.2454	0.9112	1.3809	0.9091	0.8385
75%	0.2470	0.2244	0.1921	0.1542	0.5160	0.7101	0.3869	0.4169
Median	0.0420	0.0568	0.0799	0.0475	0.1268	0.1780	0.1916	0.1180
25%	-0.0846	-0.0803	-0.0171	-0.0710	-0.1329	-0.1904	-0.0203	-0.2188
10%	-0.2238	-0.3165	-0.1024	-0.1746	-0.4146	-0.7009	-0.2179	-0.5486
5%	-0.3202	-0.4830	-0.2352	-0.2416	-0.6515	-1.2410	-0.3694	-1.1680
1%	-0.5817	-0.6943	-0.3917	-0.3246	-1.3848	-1.8933	-0.7715	-2.1845
Min	-0.7454	-1.0928	-0.5519	-0.4933	-2.1561	-2.0388	-0.7979	-2.1979

	Panel B: Performance Proportions				
		LCG	SCG	LCV	SCV
Performance	$\%\alpha_{MF,s} > 0$	59.40	62.75	72.49	57.84
Sign	$\%\alpha_{MF,s} < 0$	40.60	37.25	27.51	42.16

Table 4.7: One-Style Clientele Alphas for Mixed Mutual Funds

Table 4.7 shows statistics on the cross-sectional distribution of monthly SDF alphas for mixed funds using the one-style clientele SDFs. The one-style clientele SDFs are estimated from NAV-weighted style portfolios of small-cap funds (Small), large-cap funds (Large), value funds (Value) and growth funds (Growth). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$). The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

	Panel A: Performance and t-statistics							
	Performance				t-statistics			
	Small	Large	Value	Growth	Small	Large	Value	Growth
Mean	-0.1455	0.0144	-0.1491	0.0198	-0.0818	0.0766	-0.0131	0.0886
StdDev	0.2840	0.3919	0.3458	0.3890	0.3152	0.3680	0.3329	0.3637
(t-stat)	(-2.361)	(0.171)	(-1.983)	(0.238)				
Max	0.8960	3.1660	0.7892	2.3930	1.9882	2.5303	2.4127	3.1700
99%	0.4476	1.1315	0.5094	1.0185	0.8308	1.4655	1.2540	1.3732
95%	0.2469	0.6598	0.3585	0.7263	0.3135	0.7279	0.5717	0.7468
90%	0.1615	0.4988	0.2409	0.5217	0.1855	0.4647	0.2899	0.5274
75%	0.0240	0.1753	0.0645	0.2388	0.0164	0.1907	0.0577	0.2517
Median	-0.1332	-0.0269	-0.1398	-0.0402	-0.0808	-0.0190	-0.0743	-0.0253
25%	-0.2772	-0.1985	-0.3239	-0.2136	-0.1750	-0.1051	-0.1498	-0.1134
10%	-0.4565	-0.3669	-0.5376	-0.3703	-0.3027	-0.1916	-0.2283	-0.2089
5%	-0.5982	-0.4962	-0.6973	-0.5276	-0.4232	-0.2854	-0.3128	-0.3102
1%	-1.0747	-0.9099	-1.2100	-0.9042	-1.2660	-0.5887	-0.8010	-0.5373
Min	-3.0392	-3.2257	-3.2137	-3.2503	-2.6123	-2.1913	-2.2233	-2.0432

	Panel B: Performance Proportions				
		Small	Large	Value	Growth
Performance	$\%\alpha_{MF,s} > 0$	28.20	45.89	32.11	45.56
Sign	$\%\alpha_{MF,s} < 0$	71.80	54.11	67.89	54.44

Table 4.8: Two-Style Clientele Alphas for Mixed Mutual Funds

Table 4.8 shows statistics on the cross-sectional distribution of monthly SDF alphas for mixed funds using the two-style clientele SDFs. The two-style clientele SDFs are estimated from NAV-weighted style portfolios of large-cap growth funds (LCG), small-cap growth funds (SCG), large-cap value funds (LCV) and small-cap value funds (SCV). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$). The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics								
	Performance				t-statistics			
	LCG	SCG	LCV	SCV	LCG	SCG	LCV	SCV
Mean	-0.0564	-0.1166	-0.3124	-0.1817	0.0279	-0.0355	-0.0086	-0.1192
StdDev	0.3373	0.2932	0.4674	0.2868	0.3543	0.3325	0.3928	0.2982
(t-stat)	(-0.770)	(-1.802)	(-3.018)	(-2.742)				
Max	2.5139	0.9615	1.7441	0.7529	2.1457	2.1266	4.2506	1.7298
99%	0.8247	0.5050	0.4891	0.3953	1.5483	1.1581	1.7774	0.7638
95%	0.4963	0.3093	0.3009	0.1975	0.6598	0.4256	0.5616	0.2373
90%	0.3576	0.2139	0.1795	0.1131	0.3732	0.2604	0.2366	0.1105
75%	0.0914	0.0583	0.0065	-0.0145	0.0824	0.0445	0.0041	-0.0091
Median	-0.0769	-0.1193	-0.2598	-0.1540	-0.0479	-0.0696	-0.0992	-0.0945
25%	-0.2275	-0.2532	-0.5396	-0.3029	-0.1187	-0.1537	-0.1535	-0.1933
10%	-0.3804	-0.4464	-0.8296	-0.4827	-0.2203	-0.2560	-0.2106	-0.3651
5%	-0.5547	-0.5813	-1.1026	-0.6471	-0.3309	-0.3566	-0.2667	-0.5458
1%	-0.9304	-1.0235	-1.8729	-1.0740	-0.7104	-1.1077	-0.5137	-1.2016
Min	-3.3348	-3.0367	-4.2938	-3.0263	-2.2425	-2.1353	-1.3221	-2.1400

Panel B: Performance Proportions					
		LCG	SCG	LCV	SCV
Performance	$\%\alpha_{MF,s} > 0$	36.75	31.72	25.52	23.24
Sign	$\%\alpha_{MF,s} < 0$	63.25	68.28	74.48	76.76

Table 4.9: One-Style Clientele Alphas for Other Mutual Funds

Table 4.9 shows statistics on the cross-sectional distribution of monthly SDF alphas for Other funds using the one-style clientele SDFs. The one-style clientele SDFs are estimated from NAV-weighted style portfolios of the small-cap funds (Small), large-cap funds (Large), value funds (Value) and growth funds (Growth). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$). The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

Panel A: Performance and t-statistics								
	Performance				t-statistics			
	Small	Large	Value	Growth	Small	Large	Value	Growth
Mean	0.0220	0.1995	0.0398	0.1688	0.0388	0.2809	0.0607	0.2294
StdDev	0.2749	0.3008	0.3396	0.3119	0.2552	0.4077	0.2395	0.3886
(t-stat)	(0.369)	(3.093)	(0.539)	(2.526)				
Max	0.9495	1.1208	1.1930	1.2452	1.1840	2.2216	1.1396	1.5379
99%	0.7604	0.9820	0.9750	1.0288	1.0302	1.4209	0.9797	1.5080
95%	0.5766	0.8124	0.7837	0.8828	0.5737	1.1304	0.5453	1.0895
90%	0.3894	0.6204	0.4665	0.5844	0.3265	0.8947	0.4165	0.8572
75%	0.1767	0.3689	0.1940	0.2887	0.1026	0.4696	0.1315	0.3001
Median	-0.0234	0.1477	-0.0229	0.0913	-0.0198	0.1235	-0.0150	0.0850
25%	-0.1555	0.0074	-0.1666	-0.0256	-0.0933	0.0059	-0.0901	-0.0232
10%	-0.2500	-0.1169	-0.2841	-0.1316	-0.1832	-0.0590	-0.1419	-0.0840
5%	-0.3494	-0.1771	-0.3900	-0.2357	-0.2371	-0.1175	-0.1907	-0.1251
1%	-0.6603	-0.5136	-0.8389	-0.3783	-0.5310	-0.2331	-0.2975	-0.2347
Min	-0.7339	-0.6167	-0.9912	-0.6163	-0.6044	-0.2906	-0.3608	-0.2918

Panel B: Performance Proportions					
		Small	Large	Value	Growth
Performance	$\%\alpha_{MF,s} > 0$	45.85	76.59	47.32	68.29
Sign	$\%\alpha_{MF,s} < 0$	54.15	23.41	52.68	31.71

Table 4.10: Two-Style Clientele Alphas for Other Mutual Funds

Table 4.10 shows statistics on the cross-sectional distribution of monthly SDF alphas for Other funds using the two-style clientele SDFs. The two-style clientele SDFs are estimated from NAV-weighted style portfolios of large-cap growth funds (LCG), small-cap growth funds (SCG), large-cap value funds (LCV) and small-cap value funds (SCV). Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated alphas (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B gives the proportions of estimated alphas that are positive ($\%\alpha_{MF,s} > 0$) and negative ($\%\alpha_{MF,s} < 0$). The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

	Panel A: Performance and t-statistics							
	Performance				t-statistics			
	LCG	SCG	LCV	SCV	LCG	SCG	LCV	SCV
Mean	0.1191	0.0448	-0.0672	-0.0033	0.1518	0.0540	0.0121	0.0436
StdDev	0.2738	0.2796	0.3372	0.2768	0.3097	0.2558	0.1763	0.2970
(t-stat)	(2.003)	(0.727)	(-0.900)	(-0.051)				
Max	1.0136	1.0130	0.8002	0.8929	1.5983	1.4267	0.8150	1.6260
99%	0.8198	0.7913	0.6493	0.7166	1.3977	1.0246	0.6872	1.2924
95%	0.7058	0.6142	0.4630	0.4867	0.7095	0.5838	0.3045	0.7216
90%	0.5011	0.4522	0.3748	0.3696	0.5724	0.3677	0.2230	0.3548
75%	0.2649	0.1930	0.1546	0.1653	0.2117	0.1091	0.0763	0.0958
Median	0.0661	-0.0074	-0.0723	-0.0584	0.0524	-0.0051	-0.0342	-0.0339
25%	-0.0498	-0.1416	-0.2516	-0.1715	-0.0313	-0.0832	-0.0997	-0.0985
10%	-0.1614	-0.2344	-0.4520	-0.2880	-0.0918	-0.1459	-0.1484	-0.1802
5%	-0.2236	-0.3055	-0.5962	-0.3788	-0.1282	-0.2108	-0.1805	-0.2629
1%	-0.5661	-0.6056	-1.1785	-0.7413	-0.2619	-0.4817	-0.2725	-0.4295
Min	-0.7265	-0.6747	-1.2649	-0.9742	-0.4447	-0.5062	-0.2991	-0.6472

	Panel B: Performance Proportions				
		LCG	SCG	LCV	SCV
Performance	$\%\alpha_{MF,s} > 0$	63.90	49.27	44.39	44.39
Sign	$\%\alpha_{MF,s} < 0$	36.10	50.73	55.61	55.61

Table 4.11: Value Added from Style Clientele Alphas with $\alpha_s = 0.15\%$

Table 4.11 shows statistics on the cross-sectional distributions of monthly SDF alphas for the full sample using the one-style and two-style clientele SDFs estimated with an upper performance bound for the style portfolios of 0.15%. When there are more than one style clientele alphas for a fund, the distributions consider either its minimum (Min) or maximum (Max) alphas. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated one-style and two-style clientele alphas (denoted 1-SC alphas and 2-SC alphas, respectively) (columns under Performance) and their corresponding *t*-statistics (columns under *t*-statistics). It also reports the *t*-statistics (*t*-stat) on the significance of the mean of estimated alphas. Panel B reports proportions of estimated style clientele alphas that are larger ($\%\alpha_{MF,S} > \alpha_{FF}$) and smaller ($\%\alpha_{MF,S} < \alpha_{FF}$) than the Fama-French SDF alphas. The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the *t*-statistics.

Panel A: Performance and *t*-statistics									
	Performance					*t*-statistics			
	1-SC alphas		2-SC alphas			1-SC alphas		2-SC alphas	
	Min	Max	Min	Max		Min	Max	Min	Max
Mean	-0.1815	0.1747	-0.2390	0.0742		-0.0549	0.1990	-0.0051	0.1793
StdDev	0.3176	0.3619	0.4230	0.2901		0.3402	0.4318	0.4068	0.5528
(*t*-stat)	(-2.713)	(2.291)	(-2.682)	(1.213)					
Max	1.0794	3.1660	1.1392	2.5139		2.3518	3.1700	2.6301	4.2506
99%	0.5515	1.1269	0.6099	0.8022		1.1999	1.6782	1.6256	2.0360
95%	0.2970	0.8136	0.3491	0.5355		0.4849	0.8980	0.7328	1.2565
90%	0.1735	0.6339	0.2027	0.4219		0.2017	0.6773	0.3788	0.7883
75%	-0.0040	0.3722	0.0145	0.2251		-0.0022	0.3744	0.0102	0.3503
Median	-0.1762	0.1248	-0.1966	0.0615		-0.0888	0.1542	-0.0901	0.0830
25%	-0.3354	-0.0555	-0.4213	-0.0758		-0.1541	-0.0491	-0.1501	-0.0882
10%	-0.5375	-0.1923	-0.7227	-0.2401		-0.2461	-0.1803	-0.2290	-0.2772
5%	-0.6890	-0.3076	-0.9554	-0.3471		-0.3653	-0.3138	-0.3410	-0.4646
1%	-1.0662	-0.5596	-1.7217	-0.6794		-1.0649	-0.7693	-0.9274	-1.2388
Min	-3.2503	-2.9191	-4.2938	-2.2450		-2.7732	-2.2233	-2.1979	-2.2425

Panel B: Performance Proportions					
		1-SC alphas		2-SC alphas	
		Min	Max	Min	Max
Performance	$\%\alpha_{MF,S} > \alpha_{FF}$	24.74	76.84	29.72	77.47
Sign	$\%\alpha_{MF,S} < \alpha_{FF}$	75.26	23.16	70.28	22.53

Table 4.12: Value Added from Style Clientele Alphas with $\alpha_s = 0\%$

Table 4.12 shows statistics on the cross-sectional distributions of monthly SDF alphas for the full sample using the one-style and two-style clientele SDFs estimated with an upper performance bound for the style portfolios of 0%. When there are more than one style clientele alphas for a fund, the distributions consider either its minimum (Min) or maximum (Max) alphas. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated one-style and two-style clientele alphas (denoted 1-SC alphas and 2-SC alphas, respectively) (columns under Performance) and their corresponding *t*-statistics (columns under *t*-statistics). It also reports the *t*-statistics (*t*-stat) on the significance of the mean of estimated alphas. Panel B reports proportions of estimated style clientele alphas that are larger ($\%\alpha_{MF,S} > \alpha_{FF}$) and smaller ($\%\alpha_{MF,S} < \alpha_{FF}$) than the Fama-French SDF alphas. The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the *t*-statistics.

Panel A: Performance and *t*-statistics

	Performance					*t*-statistics			
	1-SC alphas		2-SC alphas			1-SC alphas		2-SC alphas	
	Min	Max	Min	Max		Min	Max	Min	Max
Mean	-0.1879	0.0241	-0.2144	-0.0379		-0.1366	0.0057	-0.1391	-0.0714
StdDev	0.2795	0.2919	0.3187	0.2657		0.3562	0.3654	0.4146	0.5062
(*t*-stat)	(-3.192)	(0.391)	(-3.194)	(-0.676)					
Max	0.8530	2.3333	0.7685	1.8556		1.5955	2.6818	1.8431	2.4043
99%	0.4166	0.8193	0.4261	0.5991		0.7400	1.1523	0.9709	1.2921
95%	0.2071	0.5272	0.2224	0.3687		0.2225	0.5165	0.3007	0.6990
90%	0.1135	0.3667	0.1031	0.2607		0.0956	0.3524	0.1259	0.3899
75%	-0.0335	0.1745	-0.0408	0.1059		-0.0212	0.1587	-0.0277	0.1362
Median	-0.1727	-0.0036	-0.1782	-0.0448		-0.0905	-0.0049	-0.0884	-0.0544
25%	-0.3101	-0.1374	-0.3409	-0.1658		-0.1773	-0.1430	-0.1682	-0.2287
10%	-0.4901	-0.2702	-0.5607	-0.3211		-0.4040	-0.3089	-0.4492	-0.5451
5%	-0.6349	-0.3848	-0.7430	-0.4460		-0.7124	-0.4758	-0.8764	-0.8994
1%	-0.9977	-0.6738	-1.2508	-0.7611		-1.5960	-1.1743	-1.8776	-1.8776
Min	-3.1545	-3.0302	-3.9948	-2.7440		-3.5511	-2.8152	-3.2331	-3.2331

Panel B: Performance Proportions

		1-SC alphas		2-SC alphas	
		Min	Max	Min	Max
Performance	$\%\alpha_{MF,S} > \alpha_{FF}$	16.44	58.62	21.62	52.89
Sign	$\%\alpha_{MF,S} < \alpha_{FF}$	83.56	41.38	78.38	47.11

Table 4.13: Value Added from Style Clientele Alphas with $\alpha_s = 0.3\%$

Table 4.13 shows statistics on the cross-sectional distributions of monthly SDF alphas for the full sample using the one-style and two-style clientele SDFs estimated with an upper performance bound for the style portfolios of 0.3%. When there are more than one style clientele alphas for a fund, the distributions consider either its minimum (Min) or maximum (Max) alphas. Panel A provides the mean, standard deviation (StdDev) and selected percentiles of the distributions of the estimated one-style and two-style clientele alphas (denoted 1-SC alphas and 2-SC alphas, respectively) (columns under Performance) and their corresponding t-statistics (columns under t-statistics). It also reports the t-statistics (t-stat) on the significance of the mean of estimated alphas. Panel B reports proportions of estimated style clientele alphas that are larger ($\%\alpha_{MF,S} > \alpha_{FF}$) and smaller ($\%\alpha_{MF,S} < \alpha_{FF}$) than the Fama-French SDF alphas. The data (see description in table 4.1) cover the period January 1998-December 2012. All statistics are in percentage except the t-statistics.

	Panel A: Performance and t-statistics								
	Performance					t-statistics			
	1-SC alphas		2-SC alphas			1-SC alphas		2-SC alphas	
	Min	Max	Min	Max		Min	Max	Min	Max
Mean	-0.1776	0.3286	-0.2666	0.1867		0.0068	0.3883	0.1062	0.4240
StdDev	0.3739	0.4479	0.5518	0.3272		0.4161	0.5453	0.5730	0.7010
(t-stat)	(-2.254)	(3.482)	(-2.293)	(2.708)					
Max	1.4373	3.9988	1.5098	3.1723		3.2598	4.2848	4.0363	5.8856
99%	0.7143	1.5232	0.7855	1.0721		1.7479	2.3055	2.3873	2.8530
95%	0.4091	1.1125	0.4931	0.7196		0.8015	1.3792	1.3925	1.8727
90%	0.2584	0.9113	0.3306	0.5843		0.3783	0.9901	0.8052	1.2919
75%	0.0398	0.5770	0.0841	0.3493		0.0283	0.5955	0.0959	0.6628
Median	-0.1836	0.2623	-0.2167	0.1644		-0.0912	0.3220	-0.0991	0.2343
25%	-0.3705	0.0136	-0.5034	0.0041		-0.1561	0.0112	-0.1678	0.0037
10%	-0.5977	-0.1222	-0.9096	-0.1671		-0.2326	-0.1041	-0.2301	-0.1499
5%	-0.7628	-0.2465	-1.2201	-0.2795		-0.2941	-0.2036	-0.2779	-0.2845
1%	-1.2292	-0.5169	-2.3046	-0.5725		-0.7409	-0.4741	-0.6018	-0.8388
Min	-3.3617	-2.8014	-4.5929	-1.7960		-2.0818	-1.9209	-1.7761	-1.7761

|	Panel B: Performance Proportions				
		1-SC alphas		2-SC alphas	
		Min	Max	Min	Max
Performance	$\%\alpha_{MF,S} > \alpha_{FF}$	29.05	86.32	31.66	88.42
Sign	$\%\alpha_{MF,S} < \alpha_{FF}$	70.95	13.68	68.34	11.58

Figure 4.1: Mutual Fund Style Box
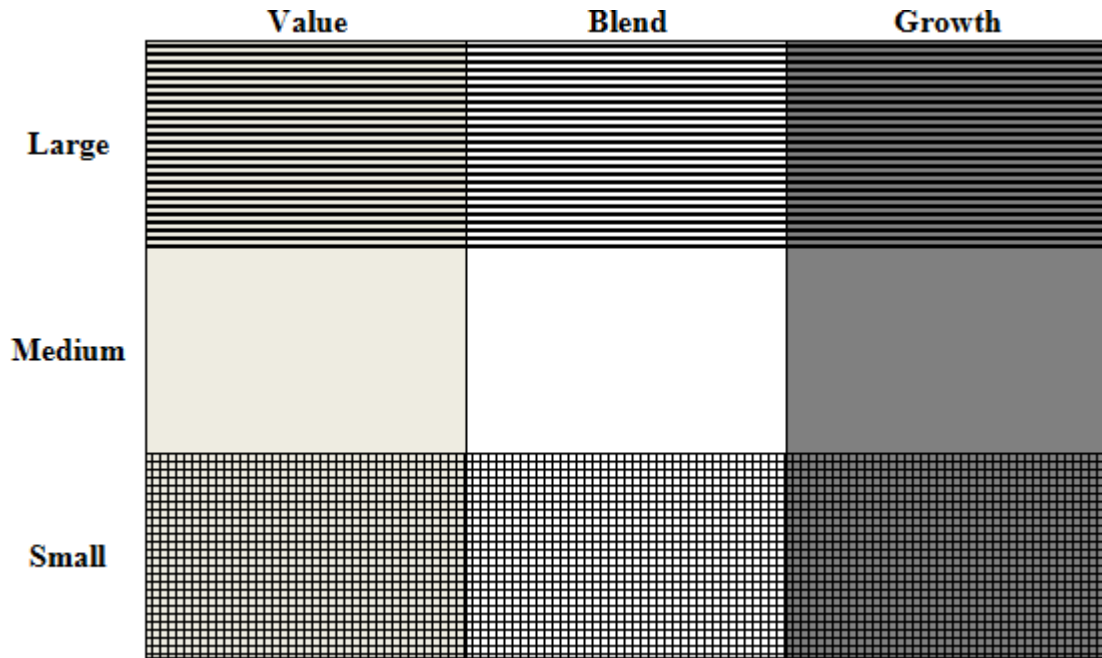
Notes: Figure 1 presents a mutual fund style box to illustrate the classification of funds. Corner funds represent funds categorized according to two style focuses and are located in the four hatched-colored boxes. They include large-cap value funds (box in light gray with horizontal lines), large-cap growth funds (box in dark gray with horizontal lines), small-cap value funds (box in light gray with squared lines) and small-cap growth funds (box in dark gray with squared lines). SLVG funds represent funds categorized according to one style focus by combining corner funds with a common style. They include small-cap funds (corner boxes with squared lines), large-cap funds (corner boxes with horizontal lines), value funds (corner boxes in light gray) and growth funds (corner boxes in dark gray). Mixed funds represent funds that are not corner funds (boxes in the medium row and in the blend column). Other funds represent funds with a style that does not fit into the style box.

Figure 4.2: One and Two-Style Clientele Stochastic Discount Factors
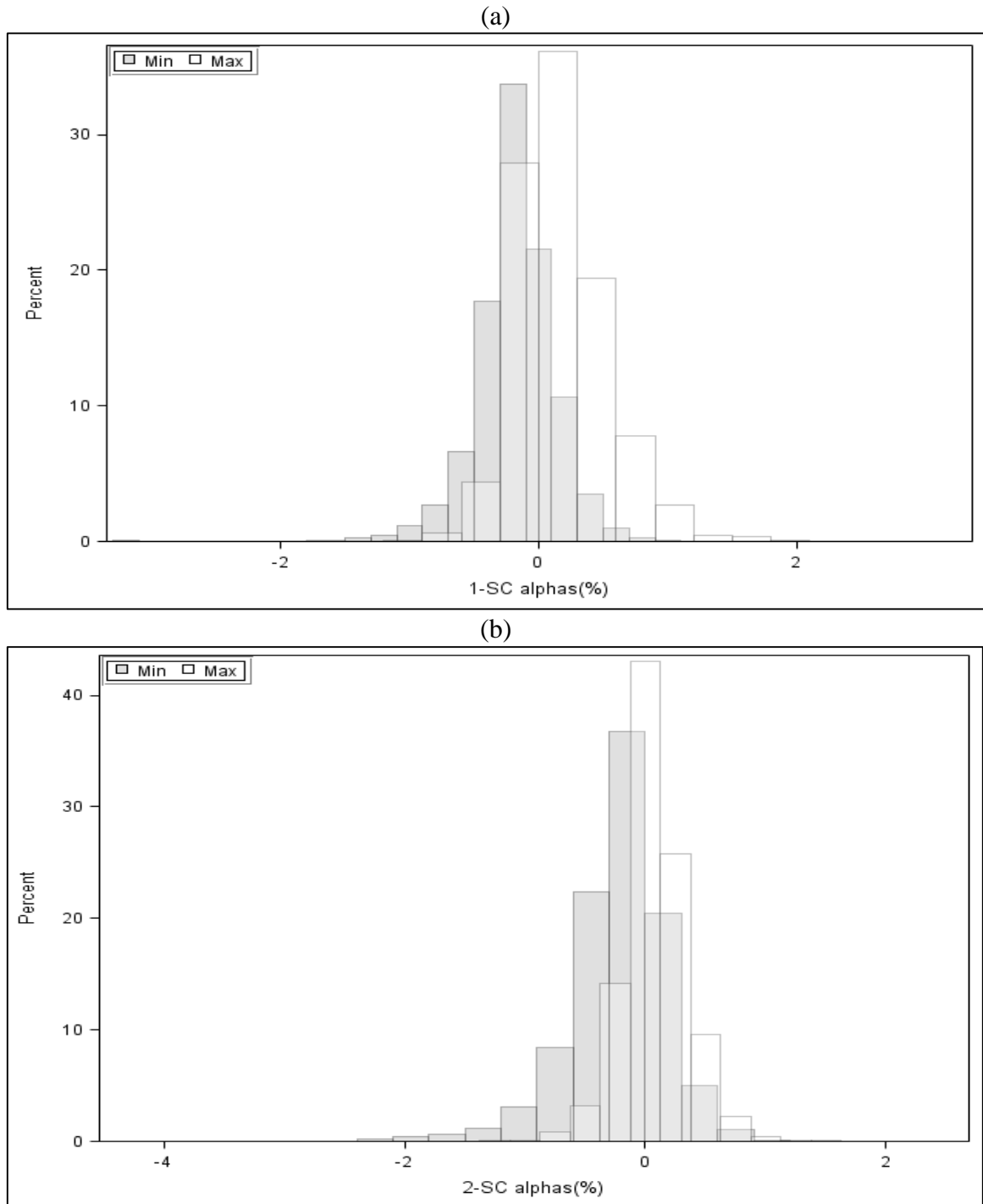
(a)
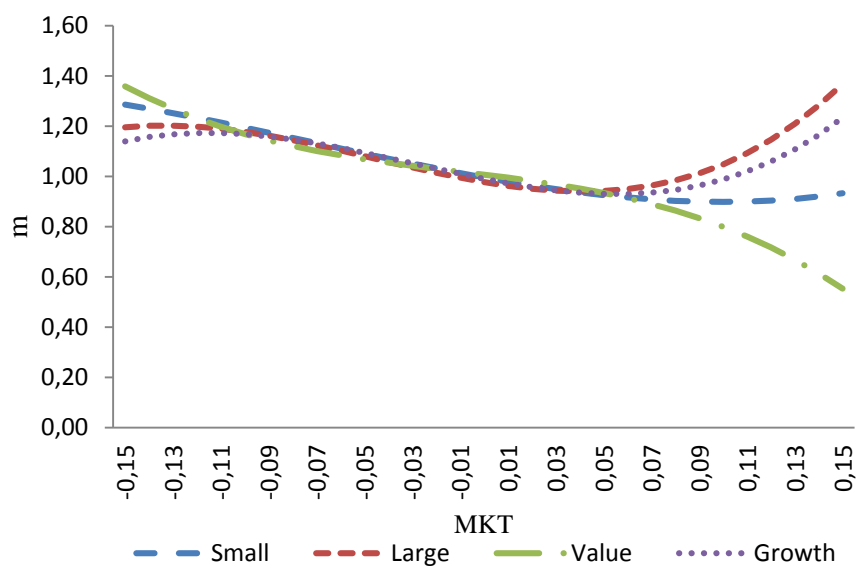
(b)
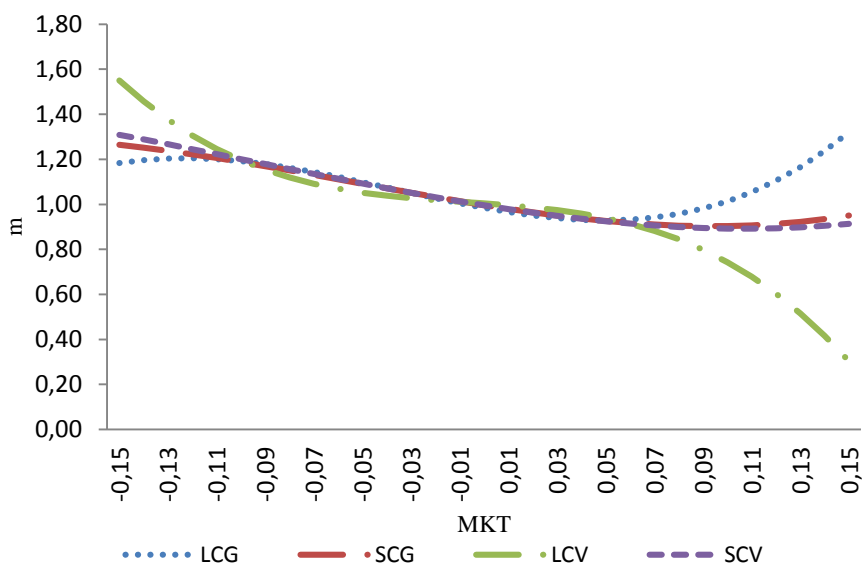
Notes: Figure 2 illustrates the stochastic discount factors (m) as a function of market portfolio returns (MKT). Figure 2a shows the stochastic discount factors for one-style clienteles. Figure 2b shows the stochastic discount factors for two-style clienteles.

Figure 4.3: Histograms of Style Clientele Alphas

(a)

(b)

Notes: Figure 3 presents histograms illustrating the distributions of the minimum (Min) and maximum (Max) style clientele alpha estimates using the one-style clientele (1-SC) SDFs (figure 3a) and the two-style clientele (2-SC) SDFs (figure 3b).

5   Conclusion

This thesis examines mutual fund performance evaluation with investor disagreement and clientele effects. We study the performance from the point of view of the best potential clienteles of mutual funds, in the sense that they value the funds at an upper performance bound. Thus, we avoid relying on the point of view of representative investors, as most of the literature has done, to instead focus on the most worthy clienteles that mutual funds could target. We combine the asset pricing bound literature with the stochastic discount factor (SDF) performance evaluation approach first proposed by Glosten and Jagannathan (1994) and Chen and Knez (1996) to develop this new measure, called the "best clientele alpha". We consider a general setup where the market is incomplete and preferences are potentially heterogeneous. Following Cochrane and Saá-Requejo (2000), we restrict the set of SDFs in an economically meaningful way using two conditions: the law-of-one-price (LOP) condition of Hansen and Jagannathan (1991) and a no-good-deal condition that rules out investment opportunities with unreasonably high Sharpe ratios.

This thesis contributes to the existing literature on the mutual fund performance evaluation with three essays. After developing, implementing and extensively checking the robustness of our performance evaluation approach in essay 1, we adapt it to answer different research questions in the remaining two essays. In essay 2, we use the best clientele alpha to diagnose the appropriateness of candidate performance measures and examine their disagreement. In essay 3, we develop clientele-specific performance measures based on the style preferences of mutual fund investors.

In the first essay, we develop a performance measure that considers the best potential clienteles of mutual funds in incomplete market with investor disagreement. Based on the law-of-one-price and no-good-deal conditions, we obtain an upper bound on admissible performance measures that identifies the most favorable evaluation. Empirically, we find that an increase in admissible investment opportunities equivalent to half the market Sharpe ratio leads to generally positive performance for best clienteles. Augmenting monthly Sharpe ratio opportunities by only 0.04 (approximately one third of the market Sharpe ratio) is sufficient for their evaluation to achieve zero alpha. Although

the literature depicts the maximum Sharpe ratio as a subjective choice, we find that a reasonably small disagreement among investors is enough to generate a positive performance from the best potential clienteles of a majority of funds.

The results of essay 1 are robust to the use of different sets of passive portfolios, conditioning information, simulated finite sample distributions for inference purposes, and adjustments for false discoveries. We also explore a conditional version of the best clientele alpha and confirm the findings of Moskowitz (2000), Kosowski (2011) and Glode (2011) that best clientele alpha estimates are more positive in recessions than in expansions. Finally, we estimate total performance disagreement defined as the difference between best and worst clientele alphas. Reinforcing the analysis of Ferson and Lin (2014), we show that investor disagreement is economically and statistically significant and can change the average evaluation of mutual funds from negative to positive, depending on the clienteles.

The best clientele performance measure gives an answer to the challenge of Ferson (2010, p.229) who argues that it is important « to identify and characterize meaningful investor clienteles and to develop performance measures specific to the clienteles ». We provide new evidence that the most favorable clienteles generally find positive values to mutual funds, in contrast to representative investors in standard performance measures that do not consider investor disagreement. Our results shed new light on the puzzling relation between the growing importance of the mutual fund industry and the negative value added by active management documented empirically, as highlighted by Gruber (1996). They also add to the findings of Ferson and Lin (2014) on the importance of heterogeneous preferences in performance evaluation.

In the second essay, we assess the validity of candidate performance measures by comparing their alphas with the alpha from the best clientele performance measure. In this comparison of the performance for representative investors versus best clienteles, we thus use the best clientele performance measure as a diagnostic tool for standard performance measures. On one hand, an inadmissibility problem occurs when a candidate alpha is greater than the upper admissible bound that is the best clientele alpha. Such analysis, in the context of performance evaluation, is similar in purpose to the investigation of Hansen and Jagannathan (1991), who propose a SDF volatility bound to diagnose candidate models in

190

the context of asset pricing. On the other hand, a misrepresentation problem occurs when a candidate alpha is lower than the best clientele alpha, because the candidate alpha provides a severe or pessimistic evaluation of the fund that is an unrealistic value for its targeted clienteles. The misrepresentation problem indicates large investor disagreement in performance evaluation.

Our diagnostic tool is uniquely positioned and complementary to the existing literature that examines the validity of performance measures with simulations (see Kothari and Warner (2001), Farnworth, Ferson, Jackson and Todd (2002), Kosowski, Timmerman, Wermers and White (2006) and Coles, Daniel and Nardari (2006), among others). Empirically, we perform the diagnosis for twelve candidate models with a total of 21 different specifications. Our implementation of the manipulation proof performance measure (MPPM) of Goetzmann, Ingersoll, Spiegel and Welch (2007) contributes to the literature because it provides monthly effective MPPM alpha estimates and their standard errors for a large sample of equity mutual funds. To our knowledge, we are the first to propose an estimation strategy that allows statistical inferences on the significance of the MPPM performance values.

The results of essay 2 show that the CAPM, the models of Fama and French (1993), Carhart (1997) and Ferson and Schadt (1996)), conditional versions of these four models, the LOP measure and the MPPM with a high risk aversion parameter misrepresent the performance evaluation of mutual funds for their best clienteles. They tend to give severe but admissible alphas that imply significant disagreement values comparable to those in Ferson and Lin (2014). In contrast, the power and habit-formation consumption-based models suffer from the inadmissibility problem as their alphas are too high. Among all models, the MPPM with a low risk aversion parameter is the most appropriate in providing admissible alphas that reflect the value of funds for their most favorable clienteles. Our novel estimation strategy for the MPPM also leads to interesting empirical findings. We find that MPPM alphas are relatively sensitive to the choice of risk aversion parameter and are difficult to estimate with statistical precision. Compared to other measures, the MPPM presents larger cross-sectional standard deviations of fund alphas.

In the third essay, as another application of our measurement approach, we focus on equity funds grouped by their investment style (i.e., value, growth, small-cap and large-cap). We develop clientele-specific performance measures to investigate fund style as an identifying attribute of meaningful investor clienteles. Such investigation is relevant as mutual funds cater to and attract specific clienteles through their widely publicized investment style. We propose a performance framework with investor disagreement that allows for the identification of meaningful SDFs for style clienteles. To implement the framework, we use a new style classification method to better exploit existing objective code data from Lipper and account for code changes and missing codes.

Empirically, we explore the economic properties of the marginal preferences reflected in the SDFs for style clienteles. The results show that the preferences implied by the SDFs have similar risk aversion but differ in their behavioral features in extreme market states. We also examine a clientele-specific performance evaluation using the SDFs for style clienteles. The evaluation finds that funds assigned to equity styles have a neutral to positive performance when they are evaluated with their relevant clientele-specific measure. The performance of the other funds is sensitive to the clienteles, and in particular their behavioral tendencies for optimism or pessimism. We finally study the value added of the mutual fund industry from the perspective of different style clienteles. We find that the sign of the value added is ambiguous and depends on the choice of measures.

The results of essay 3 document the performance disagreement of investors attracted to different fund styles by showing that preferences and alphas differ for size and value mutual fund clienteles. They support the conjecture of Ferson (2010) that properly evaluating mutual funds might require clientele-specific measures based on meaningful investor clienteles.

Overall, the findings of the three essays provide numerous contributions to the literatures associated with mutual funds. We improve on existing performance evaluation approaches by developing a performance measure with investor disagreement that focuses on the class of investors most favorable to mutual funds (i.e., the best clientele alpha), by estimating the MPPM with a new strategy that allows statistical inferences on the significance of the performance values, and by proposing clientele-specific performance

measures based on the style preferences of mutual fund investors. We contribute to the literature on the value added by active management by robustly showing that best clienteles generally evaluate positively the performance of mutual funds, and by finding that the value added is different for various size and value style clienteles.

We expand on the literature on the benchmark choice problem by documenting that many standard performance evaluation measures suffer from either an inadmissibility problem or a misrepresentation problem. We also ensure that the proposed best clientele measure and style-clientele-specific measures do not suffer from benchmark choice problem by imposing that they correctly price passive portfolios. We complement the literature on the comparison of performance measures with simulations by offering a diagnostic tool that is uniquely positioned to assess the validity of performance measures because it is based on admissible bounds instead of simulations, and by showing that the tool is useful to empirically identify problems in existing measures.

We contribute to the literature on heterogeneous preferences and investor disagreement by finding that investor disagreement can be significant enough to change the evaluation of funds from positive to negative, by showing that the disagreement between the representative investors implicit in many standard measures and best clienteles is significant and reinforces the analysis of Ferson and Lin (2014), and by documenting that behavioral characteristics and performance values differ for size and value style clienteles. Finally, we add to the literature on mutual fund styles by showing that equity styles are relevant to identify meaningful investor clienteles and by introducing a new method to better exploit available investment objective code data and account for their stability and quality.

Our research on mutual fund performance evaluation can be extended in numerous ways. First, other conditions can be imposed to restrict the set of SDFs for performance evaluation in an economically meaningful way, like the maximum-gain-loss-ratio condition of Bernardo and Ledoit (2000). These conditions can potentially adapt our approach to the performance measurement of portfolios with nonlinear payoffs, such as hedge funds. Second, we can use the best clientele alpha as a diagnostic tool to document the

misrepresentation and inadmissibility problems of numerous other performance models. Third, we can study the determinants of investor disagreement by looking at fund characteristics, like expense ratios and turnovers, taxes and transaction costs. Finally, future research could investigate the impact of investor heterogeneity and disagreement on mutual fund flows to better understand the demands of investor clienteles. Ultimately, we agree with Ferson (2010) and Ferson and Lin (2014) on their calls for more research on clientele effects in performance evaluation.

6 References

Bernardo, A. E., Ledoit, O., (2000), "Gain, loss, and asset pricing", *Journal of Political Economy*, 108, 144-172.

Carhart, M. M., (1997), "On persistence in mutual fund performance", *Journal of Finance*, 52, 57-82.

Chen, Z., Knez, P. J., (1996), "Portfolio performance measurement: theory and applications", *Review of Financial Studies*, 9, 511-555.

Cochrane, J., Saá-Requejo, J., (2000), "Beyond arbitrage: good-deal asset price bounds in incomplete markets", *Journal of Political Economy*, 108, 79-119.

Coles, J. L., Daniel, N. D., Nardari, F., (2006), "Does the misspecification of timing strategy or timing index affect inference in measuring mutual fund performance?", Unpublished Working Paper, Arizona State University.

Fama, E. F., French, K. R., (1993), "Common risk factors in the returns on stocks and bouds", *Journal of Financial Economics*, 33, 3-56.

Farnsworth, H., Ferson, W., Jackson, D., Todd, S., (2002), "Performance evaluation with stochastic discount factors", *Journal of Business*, 75, 473-503.

Ferson, W. E., (2010), "Investment performance evaluation", *Annual Review of Financial Economics*, 2, 207-234.

Ferson, W. E., Lin, J., (2014), "Alpha and performance measurement: the effect of investor heterogeneity", *Journal of Finance*, 69, 1565-1596.

Ferson, W. E., Schadt, R., (1996), "Measuring fund strategy and performance in changing economic conditions", *Journal of Finance*, 51, 425-462.

Glode, V., (2011), "Why mutual funds underperform", *Journal of Financial Economics*, 93, 546-559.

Glosten, L., Jagannathan, R., (1994), "A contingent claims approach to performance evaluation", *Journal of Empirical Finance*, 1, 133-166.

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

Gruber, M. J., (1996), "Another puzzle: The growth in actively managed mutual funds", *Journal of Finance*, 51, 783-810.

Hansen, L. P., Jagannathan, R., (1991), "Implication of security market data for models of dynamic economies", *Journal of Political Economy*, 99, 225-262.

Kosowski, R., (2011), "Do mutual funds perform when it matters most to investors? U.S. mutual fund performance and risk in recessions and expansions", *Quarterly Journal of Finance*, 1, 607-664.

Kosowski, R., Timmermann, A., Wermers, R., White, H. L., (2006), "Can mutual fund "stars" really pick stocks? Evidence from a bootstrap analysis", *Journal of Finance*, 61, 2251-2295.

Kothari, S. P., Warner, J. B., (2001), "Evaluating mutual fund performance", *Journal of Finance*, 56, 1985-2010.

Moskowitz, T., (2000), "Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses: Discussion", *Journal of Finance*, 55, 1695-1703.