2019

Essays on asset allocation and delegated portfolio management

BOSTON UNIVERSITY

QUESTROM SCHOOL OF BUSINESS

Dissertation

ESSAYS ON ASSET ALLOCATION AND DELEGATED PORTFOLIO MANAGEMENT

by

QIAOZHI HU

B.Econ., Nankai University, 2012
M.Sc., Rutgers University, 2014

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2019

Approved by

First Reader

Marcel Rindisbacher, Ph.D.
Senior Associate Dean for Faculty & Research
Professor of Finance

Second Reader

Jérôme B. Detemple, Ph.D.
Morton H. and Charlotte Friedman Professor in Management
Professor of Finance

Third Reader

Andrea M. Buffa, Ph.D.
Assistant Professor of Finance

Acknowledgments

I would like to express my deepest appreciation to my advisors, Professor Marcel Rindisbacher and Professor Jérôme Detemple, for their continuous support, encouragement and invaluable advice during my entire Ph.D. study and research. Their patience and trust allowed me to conduct research in an excellent environment. I have benefitted tremendously from their extensive knowledge and constructive criticism. I am also very grateful to Professor Andrea Buffa for his insightful suggestions and comments through the writing process of this dissertation.

I must also thank Andrew Lyasoff, Dirk Hackbarth, Eric Jacquier, Gustavo Schwenkler, Rodolfo Prieto, Scott Robertson and all other professors at Boston University who gave me a solid foundation in mathematical finance. Thanks should also go to my fellow graduate students Huangyu Chen, Li Chen, Diogo Duarte, Jianing Duan, Liang Dai, Kyounghwan Lee, Ketong Lin, Qing Liu, Dapeng Shang, Yang Wang, Liangliang Zhang and Hannan Zheng for the interesting discussions on finance topics.

Finally, I am deeply indebted to my family. I thank my father Kongliang Hu and mother Lianfang Liu for their care, understanding, and love. Special thanks to my wife Yue Wu for her trust, encouragement, and for sticking with me through many ups and downs, without which I cannot stand where I am today.

ESSAYS ON ASSET ALLOCATION AND DELEGATED PORTFOLIO MANAGEMENT

QIAOZHI HU

Boston University, Questrom School of Business, 2019

Major Professors: Marcel Rindisbacher, Ph.D.
Senior Associate Dean for Faculty & Research
Professor of Finance

Jérôme B. Detemple, Ph.D.
Morton H. and Charlotte Friedman Professor in Management
Professor of Finance

ABSTRACT

Asset allocation and portfolio decisions are at the heart of money management and draw great attention from both academics and practitioners. In addition, the segmentation of fund investors (i.e., the clientele effect) in the money management industry is well known but poorly understood. The objective of this dissertation is to study the implications of regime switching behaviors in asset returns on asset allocation and to analyze the clientele effect as well as the impact of portfolio management contracts on fund investment.

Chapter 2 presents an innovative regime switching multi-factor model accounting for the different regime switching behaviors in the systematic and idiosyncratic components of asset returns. A Gibbs sampling approach for estimation is proposed to deal with the computational challenges that arise from a large number of assets and multiple Markov chains. In the empirical analysis, the model is applied to study sector

exchange-traded funds (ETFs). The idiosyncratic volatilities of different sector ETFs exhibit a strong degree of covariation and state-dependent patterns, which are different from the dynamics of their systematic component. In a dynamic asset allocation problem, the certainty equivalent return is computed and compared across various models for an investor with constant relative risk aversion. The out-of-sample asset allocation experiments show that the new regime switching model statistically significantly outperformed the linear multi-factor model and conventional regime switching models driven by a common Markov chain. The results suggest that it is not only important to account for regimes in portfolio decisions, but correct specification about the structure and number of regimes is of equal importance.

Chapter 3 proposes a rational explanation for the existence of clientele effects under commonly used portfolio management contracts. It shows that although a fund manager always benefits from his market timing skill, which comes from his private information about future market returns, the value of the manager's private information to an investor can be negative when the investor is sufficiently more risk-averse than the manager. This suggests different clienteles for skilled and unskilled funds. Investors in skilled funds are uniformly more risk-tolerant than investors in unskilled funds. Moreover, a comparative statics analysis is conducted to investigate the effects of the manager's skill level, contract parameters, and market conditions on an investor's fund choice. The results suggest that the investors who are sufficiently more risk-averse than the manager should include fulcrum fees in the contract to benefit from the skilled manager's information advantage.

Contents

List of Tables

List of Figures

List of Abbreviations

AUM	....	Asset Under Management
CDF	....	Cumulative Distribution Function
CER	....	Certainty Equivalent Return
CRRA	....	Constant Relative Risk Aversion
CRSP	....	Center for Research in Security Prices
ETF	....	Exchange-traded Funds
ICI	....	Investment Company Institute
PIPR	....	Private Information Price of Risk
RMSE	....	Root Mean Square Error
RRA	....	Relative Risk Aversion
RRP	....	Relative Risk Prudence
SEC	....	Securities and Exchange Commission
SPD	....	Stochastic Price Density
SSD	....	Second-order Stochastically Dominate

Chapter 1

Introduction

The ongoing U.S. stock bull market since March 9, 2009 has already become the longest one in American history. But the remarkable run and economic expansion will not last forever due to the cyclical nature of economics. Financial markets often exhibit dramatic changes in their patterns, associated with economic events or unexpected changes in economic policy. While some changes are permanent breaks, often the asset returns preserve the changed patterns for a period of time and switch back to original or other regimes. Many studies have documented the regime switching behaviors not only in stock returns but also in systematic pricing factors (Ang and Bekaert, 2002; Ang and Chen, 2002; Guidolin and Timmermann, 2007b, 2008; Turner et al., 1989). Therefore, it is a critical task to investigate the implications of regimes switching on asset allocation and portfolio decisions.

There is a growing body of literature that highlights the importance of accounting for regime switching in asset allocation (Ang and Bekaert, 2004; Guidolin and Timmermann, 2007a; Tu, 2010). However, most of the existing literature on regime switching assumes that the regime shift in all assert returns is governed by a common state variable.[1] In addition, there are no studies that examine the implication of regime switching idiosyncratic volatilities on asset allocation. Furthermore, it is usually difficult to apply regime switching models to study a large number of assets due to the curse of dimensionality.

[1]One exception is Vial (2015), in which the author considers the time-varying synchronization of regimes in different countries' gross total return indices.

To address these limitations and challenges, Chapter 2 presents a Markov regime switching multi-factor model accounting for different regime switching behaviors in the systematic and idiosyncratic components of asset returns. A Bayesian approach is employed to deal with the computational challenge arising from estimating a high-dimensional regime switching model. In the empirical analysis, this approach is applied to study nine sector ETFs and a regime switching version of Fama-French three-factor model is considered. Linear multi-factor model is strongly rejected by the data. The systematic component of the sector ETFs exhibits three different regimes, which are characterized by the difference in the risk premia, factor loadings, and variance-covariance matrix. The three regimes can be interpreted as bull, slow-growth, and bear state. While two regimes in the dynamics of sector ETFs' idiosyncratic volatilities are identified and they have natural interpretations as low idiosyncratic volatility and high idiosyncratic volatility state.

A dynamic asset allocation problem is studied under this model. The results show that optimal asset allocation weights vary significantly across different regimes. The out-of-sample asset allocation experiment confirms that the economic gain of accounting for regime switching in investment is significant. In addition, the new regime switching model delivers superior risk-adjusted investment performance compared to regime switching models with a common Markov chain. The outperformance is measured in terms of certainty-equivalent returns, ranging from 1% to 5% depending on the investment horizon. The economic gain or outperformance is more significant for investors with a shorter investment horizon.

Standard portfolio choice theory ignores the delegated portfolio managers (e.g., mutual fund managers) and assumes that investors are directly managing their own money. This assumption has its root in the equity market in the mid-1950s but it no longer does: more than 100 million US retail investors accounted for half of all US households delegate the management of their wealth to professional portfolio managers at the year-end 2018 according to a report by Investment Company Institute (ICI).[2] One interesting phenomenon in delegated portfolio management is the clientele effect that fund investors appear to segment the markets. For example, Blackburn et al. (2009) find that there are different investor clienteles in value and growth funds and risk aversion is an important attribute to differentiate these two groups of investors. Investors in value funds appear to be more risk-averse than investors in growth funds. In addition, Chan et al. (2002) find that growth managers have greater abilities to generate alpha than value managers. The existing literature attributes the clientele effects to irrationality, investor sophistication or other psychological tendencies (e.g., Barberis and Shleifer, 2003; Del Guercio and Reuter, 2014).

In contrast to the behavioral attributes that may or may not determine fund clienteles, Chapter 3 presents a rational model for the clientele effect and derives the emergence of clienteles endogenously based purely on risk aversion.[3] In this model, there is heterogeneity in the fund manager's skill which comes from privately informed information about the future market returns whose content is unknown to investors. The anticipative information is always valuable to the manager and increases in the information precision (i.e. skill). However, when the manager and the investors exhibit different risk preferences, the private nature of this information can be costly and adverse to the investors. Investors whose risk aversion lies above a threshold value that depends on the fund manager's risk prudence would prefer to choose the unskilled fund rather than skilled ones. Thus, the fund investor clientele is endogenously segmented. The relatively risk-tolerant investors will prefer skilled funds, whereas the highly risk-averse investors will prefer unskilled funds. This result

[2] `https://www.ici.org/pdf/2019_factbook.pdf`

[3] Chapter 3 is based on my co-authored paper Hu and Rindisbacher (2018). Some contents of the paper are also incorporated into the introductory and conclusive chapters.

thus provides theoretical justification for some recent empirical findings of the clientele effect in the mutual fund industry (e.g., Del Guercio and Reuter, 2014).

Another key feature of delegated portfolio management is the presence of portfolio management contracts between fund managers and investors. Chapter 3 also examines the impacts of commonly used contracts on the investment decisions of fund managers and investors as well as the clientele effect. In the absence of performance fees, investors whose relative risk aversion larger than the relative prudence coefficient of the logarithmic fund managers always prefer the unskilled fund irrespective of the skilled funds' skill level. While in the presence of symmetric or asymmetric performance fees, the relative risk aversion threshold that separates the clienteles is affected by the skill level, contract parameters, and market conditions. The result shows that including a symmetric fee in the manager's compensation contract could lead to a higher value of manager's information to sufficiently risk-averse investors than that under option-like asymmetric fees. This suggests that investors who are sufficiently more risk-averse than the manager may choose to include a symmetric performance fee in the portfolio management contract to better exploit the manager's private information.

Chapter 2

A Markov Regime Switching Model for Asset Allocation

2.1 Introduction

Conventional asset pricing studies use linear factor models with stable coefficients to examine asset returns and assume that the distribution of returns and pricing factors does not change over time. However, financial time series often change their pattern and the abrupt changes can be recurring (e.g., bull versus bear market). There is substantial empirical evidence of multiple regimes in the distribution of asset returns (Ang and Bekaert, 2002; Ang and Chen, 2002; Guidolin and Timmermann, 2008; Turner et al., 1989). Guidolin and Timmermann (2007b) find strong evidence of state-dependent variations in the joint distribution of returns on market, size, and value factor portfolios. The risk premia, volatility, and correlations between these factor portfolios are found to be driven by an underlying state variable. This chapter further extends the findings to a Markov switching multifactor model and accounts for the different regime in the idiosyncratic component of asset returns. Such an extension significantly improves the fit of factor models and yields superior portfolio performance as explained next.

Our study first documents that the cyclical variations of the idiosyncratic volatility at the sector level and idiosyncratic volatility series possess substantial common variations across sectors. These two regimes in the idiosyncratic returns of sector ETFs have natural interpretations as the low-volatility and the high-volatility regime. In addition, we find that the regimes identified from the idiosyncratic returns are different from the regimes identified from systematic components of sector ETFs, which are characterized as the bull, slow-growth, and bear states. This motivates the introduction of a new regime switching model by decomposing asset returns into idiosyncratic and systematic components and modeling the two different cycles using two different Markov chain latent variables. The out-of-sample asset allocation experiment shows that the economic gains of accounting for regimes in the asset returns are significant and the proposed regime switching model statistically significantly outperformed traditional regime switching models with a common state variable.

Markov regime switching models have gained popularity in finance and economic literature since the seminal work of Hamilton (1989). When applied to equity returns, regime switching models are well known for their ability to capture many stylized facts such as fat tails, volatility clustering, skewness, autocorrelation, and time-varying correlations. There are also good economic reasons behind these model and the empirical findings are compatible with equilibrium models. Whitelaw (2000) considers a general equilibrium exchange economy with regime shifts in the consumption growth and the conditional moments of stock returns vary considerably across different states. Ang and Timmermann (2012) introduce a simple equilibrium model with a regime switching dividend or consumption growth and show that strong dependence in expected returns and volatility are endogenously generated.

There is a growing body of literature that studies the implication of regime switching on asset allocation. For example, Ang and Bekaert (2004) consider the presence of two regimes in expected equity returns and volatility, and they find that substantial economic value is added by incorporating regimes into asset allocation. The outperformance of the regime switching strategy is largely due to its market timing ability.

Our study shares some intuition with this paper and extends Ang and Bekaert's by considering a more realistic assumption that investors need to infer the unobservable regimes. (Guidolin and Timmermann, 2007a) study the asset allocation implications of the regimes in the joint distribution of stock and bond returns. The optimal portfolio weights vary significantly across the four regimes as investors revise their beliefs about the state variables. Our study focuses on a different asset allocation problem to sector ETFs and T-bill. Furthermore, we employ a new Markov switching framework and Bayesian approach for estimation, which makes it possible to deal with high-dimensional Markov switching models that are very difficult to handle using maximum likelihood estimation (MLE). Tu (2010) also uses a Bayesian framework to study a static portfolio choice problem that takes regime switching into account. He finds that the economic gains of incorporating regimes are significant irrespective of concerns about model or parameter uncertainty. In contrast to Tu's model with a common Markov chain, asset returns in our study are driven by two different latent state variables in their idiosyncratic and systematic components, which reflects the fact the two components have different regime cycles. In addition, we consider the effects of rebalancing on asset allocation by studying a dynamic problem.

This study is also closely related to a large body of literature that focuses on idiosyncratic volatility, defined as the residual standard deviation in factor models. It is documented that the idiosyncratic volatility exhibited cyclical variations (e.g., Brandt et al., 2010; Bartram et al., 2016). Bekaert et al. (2012) find that idiosyncratic volatility can be fitted using a regime-switching autoregressive process with occasional shifts to a low-volatility or high-volatility regime. Explanations such as fundamental cash flow variability, business cycle effects, and market aggregate volatility can only partially explain the regime-switching behavior in idiosyncratic volatility. They argue that the phenomenon might be related to undiversifiable tail risk in crisis periods and it is hard to capture by a linear model. Our analysis shares some intuition and finds that the non-linear patterns in the idiosyncratic volatility can be well captured by the new regime switching model.

Another strand of idiosyncratic volatility literature has found that firm-level and industry-level idiosyncratic volatilities possess a high degree of co-movement and argue that idiosyncratic volatility is a common pricing factor (e.g., Ang et al., 2006, 2009). Herskovic et al. (2016) argue that idiosyncratic return residuals are priced and not fully diversifiable due to the presence of non-traded assets, such as human capital. They develop an incomplete market model with heterogeneous investors whose consumption growth shares the same volatility structure as firms' idiosyncratic cash flow growth. The common idiosyncratic volatility factor becomes the driver of stock return volatility and dispersion in household earnings growth. A related paper Franco et al. (2017) add the volatility factor to the standard Fama-French three-factor model and find increasing the number of factors will impede the model's interpretability due to the fact that the volatility factor is highly negatively correlated to the market and size factor. As an alternative approach, they suggest considering a Markov-switching framework to improve the explanatory power of the linear multi-factor models. In addition, we find that adding a volatility factor will not eliminate the regime switching behavior in the standard deviation of residuals in the extended factor model.

The rest of the chapter is structured as follows. Section 2.2 introduces the new regime switching model. Section 2.3 describes the Gibbs sampling approach to estimating the regime switching parameters and state probabilities of multiple Markov chains. Section 2.4 conducts an empirical analysis of regimes in the factor portfolios and idiosyncratic returns of sector ETFs as well as their economic interpretations. Section 2.5 sets up the investor's asset allocation problem and Section 2.6 computes the optimal portfolio weights and presents the asset allocation results. Section 2.7

conducts a robustness analysis. Conclusions are in Section 2.8.

2.2 A Regime Switching Multi-factor Model

To capture the nonlinear effects in the joint distribution of asset returns and factors, consider a $K \times 1$ vector of benchmark factors, $\boldsymbol{F}_t = (F_{1t}, \ldots, F_{Kt})$ and an $N \times 1$ vector of asset returns in excess of the constant T-bill rate, $\boldsymbol{R}_t = (R_{1t}, \ldots, R_{Nt})$. Suppose that the risky asset and factor returns evolve as follows:

$$\boldsymbol{F}_t = \boldsymbol{\mu}(S_t) + \boldsymbol{\Omega}^{1/2}(S_t)\boldsymbol{\epsilon}_t, \tag{2.1}$$

$$\boldsymbol{R}_t = \boldsymbol{\alpha}(Z_t) + \boldsymbol{\beta}(S_t)\boldsymbol{F}_t + \boldsymbol{\Sigma}^{1/2}(Z_t)\boldsymbol{u}_t \tag{2.2}$$

where S_t and Z_t are two independent, discrete-time, homogenous, irreducible and ergodic first-order Markov chains with a finite number of k_s and k_z regimes and constant $k_s \times k_s$ transition probability matrix $\mathbf{P}_s$ and $k_z \times k_z$ matrix $\mathbf{P}_z$ respectively, whose (i, j)-th elements are given by

$$\mathbf{P}_s(i,j) = \Pr(S_t = i | S_{t-1} = j), \quad i, j = 1, \ldots, k_s, \tag{2.3}$$

$$\mathbf{P}_z(i,j) = \Pr(Z_t = i | Z_{t-1} = j), \quad i, j = 1, \ldots, k_z. \tag{2.4}$$

$\boldsymbol{\mu}$ represents a $K \times 1$ vector of state-dependent risk premia, $\boldsymbol{\alpha}$ is a $N \times 1$ vector of state-dependent abnormal returns, $\boldsymbol{\beta}$ is a $N \times K$ matrix of state-dependent factor loading, $(\boldsymbol{\epsilon}_t, \boldsymbol{u}_t)' \sim$ IID $N(\mathbf{0}, \boldsymbol{I}_{K+N})$. The matrix $\boldsymbol{\Omega}^{1/2}(S_t)$ represents the state-dependent Cholesky factorization of the covariance matrix of factors $\boldsymbol{F}_t$:

$$\boldsymbol{\Omega}^{1/2}(S_t)\left(\boldsymbol{\Omega}^{1/2}(S_t)\right)' = \boldsymbol{\Omega}(S_t) \equiv Var[\boldsymbol{F}_t | S_t]. \tag{2.5}$$

Similarly, $\mathbf{\Sigma}^{1/2}(Z_t)$ represents the state-dependent Cholesky factorization of the covariance matrix of "idiosyncratic" returns, $\boldsymbol{r}_t(Z_t) = \boldsymbol{\alpha}(Z_t) + \mathbf{\Sigma}^{1/2}(Z_t)\boldsymbol{u}_t$:

$$\mathbf{\Sigma}^{1/2}(Z_t)\left(\mathbf{\Sigma}^{1/2}(Z_t)\right)' = \mathbf{\Sigma}(Z_t) \equiv Var[\boldsymbol{r}_t|Z_t]. \tag{2.6}$$

The residuals $\boldsymbol{r}_t$ are relative to the factor model described in Equation 2.2, but they are not necessarily idiosyncratic or asset-specific. In Section 2.7, we consider an extensive set of factors, but it shows that the residuals still contain a latent dynamic component that drives the strong covariation of idiosyncratic volatility. Thus the idiosyncratic risk is not fully diversifiable. Possible explanations include undiversifiable tail risk exposure in crisis periods or the presence of non-traded assets (e.g., Bekaert et al., 2012; Herskovic et al., 2016). In Section 2.4 we find that the idiosyncratic returns can be well captured a two-state regime switching process with occasional shifts to a low-volatility or high-volatility regime, which is in line with the findings of Bekaert et al. (2012).

Factor portfolios $\boldsymbol{F}_t$ are driven by the state variable S_t. The state-dependent mean vector $\boldsymbol{\mu}(S_t)$ and variance-covariance matrix $\mathbf{\Omega}(S_t)$ are constant within each state. We assume that individual risky asset's factor loadings $\boldsymbol{\beta}(S_t)$ are time-varying and depend on the same Markov chain S_t in the benchmark factors.[1] This allows the systematic component of asset returns, $\boldsymbol{\beta}(S_t)\boldsymbol{F}_t(S_t)$, is driven by the same Markov chain, S_t. The other Markov chain Z_t controls the idiosyncratic part of asset returns and determines the abnormal returns $\boldsymbol{\alpha}(Z_t)$ and variance-covariance matrix $\mathbf{\Sigma}(Z_t)$ of idiosyncratic returns.

In this model, investors cannot observe the underlying state variables S_t or Z_t even at time t. The true states can only be inferred from the realizations of asset and factor portfolio returns. Thus investors face a portfolio choice problem with partial

[1] The empirical results in Section 2.4 documents that the factor loadings are statistically significantly different across regimes identified from factor returns.

information.

Existing multivariate regime-switching models usually assumes that a common regime affects all mean return vector, factor loadings, and variance-covariance matrix (e.g., Tu, 2010). However, as demonstrated in our empirical findings idiosyncratic risk does not necessarily share the same latent variable that drives the systematic risk factors. Idiosyncratic volatility usually accounts for a substantial portion of "total" volatility and is found to be related to expected returns. It is thus important to model the different dynamics of idiosyncratic and systematic parts of asset returns. The proposed Markov regime switching multi-factor model captures this idea and provides a more powerful framework to conduct performance and risk analysis.

2.3 Estimation of the Model

Most multivariate regime switching studies focus on a very small number of assets and conduct parameter estimation using the MLE approach. Due to the curse of dimensionality, it is very difficult to apply MLE to study high-dimensional regime switching problems. On the other hand, parameter uncertainty is well documented to have a strong impact on portfolio choice (e.g., Tu and Zhou, 2010; Pastor and Stambaugh, 2000). To handle the issues of high-dimensional asset returns and parameter uncertainty, we propose a Bayesian approach to estimating the new Markov regime switching model based on Gibbs sampling. The modified MCMC approach provides a novel way to jointly identify the dynamics of multiple Markov chains and model parameters.[2] Under the Bayesian framework, both the Markov-switching state variables $\{S_t\}_{t=1}^T$, $\{Z_t\}_{t=1}^T$ and the model parameters $\boldsymbol{\theta} \equiv \{\{\boldsymbol{\alpha}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{k_z}, \{\boldsymbol{\beta}_i, \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\}_{i=1}^{k_s}, \boldsymbol{P}_s, \boldsymbol{P}_z\}$ are treated as random variables given the data $\{\boldsymbol{R}_t, \boldsymbol{F}_t\}_{t=1}^T$.

A modified Hamilton (1994) filtering algorithm is used to calculate the filtered

[2] Previous studies usually identify multiple Markov chains one by one with strong assumptions on the Markov chain structure (e.g., Billio et al., 2010).

probabilities

$$\xi^s_{t|t} \equiv \begin{pmatrix} \Pr(S_t = 1|I_t; \boldsymbol{\theta}) \\ \vdots \\ \Pr(S_t = k_s|I_t; \boldsymbol{\theta}) \end{pmatrix} \quad \text{and} \quad \xi^z_{t|t} \equiv \begin{pmatrix} \Pr(Z_t = 1|I_t; \boldsymbol{\theta}) \\ \vdots \\ \Pr(Z_t = k_z|I_t; \boldsymbol{\theta}) \end{pmatrix},$$

where I_t denotes available information at time t, namely realizations of asset and factor returns up to time t. The filtering algorithm proceeds in two steps: a prediction step and an updating step. In the prediction step, the predicted probability is computed according to

$$\xi^s_{t|t-1} = \boldsymbol{P}_s \xi^s_{t-1|t-1} \quad \text{and} \quad \xi^z_{t|t-1} = \boldsymbol{P}_z \xi^z_{t-1|t-1}.$$

In the update step, the information in the data at time t is used to update the predicted estimate via the formula

$$\xi^s_{t|t} = \frac{\left(\xi^s_{t|t-1}(\xi^z_{t|t-1})' \odot f(\boldsymbol{R}_t, \boldsymbol{F}_t|S_t, Z_t, I_{t-1}; \boldsymbol{\theta})\right) \mathbf{1}_{k_z}}{\mathbf{1}'_{k_s} \left(\xi^s_{t|t-1}(\xi^z_{t|t-1})' \odot f(\boldsymbol{R}_t, \boldsymbol{F}_t|S_t, Z_t, I_{t-1}; \boldsymbol{\theta})\right) \mathbf{1}_{k_z}}, \tag{2.7}$$

$$\xi^z_{t|t} = \frac{\left(\xi^s_{t|t-1}(\xi^z_{t|t-1})' \odot f(\boldsymbol{R}_t, \boldsymbol{F}_t|S_t, Z_t, I_{t-1}; \boldsymbol{\theta})\right)' \mathbf{1}_{k_s}}{\mathbf{1}'_{k_s} \left(\xi^s_{t|t-1}(\xi^z_{t|t-1})' \odot f(\boldsymbol{R}_t, \boldsymbol{F}_t|S_t, Z_t, I_{t-1}; \boldsymbol{\theta})\right) \mathbf{1}_{k_z}} \tag{2.8}$$

where $\odot$ denotes element by element multiplication and $f(\boldsymbol{R}_t, \boldsymbol{F}_t|S_t, Z_t, I_{t-1}; \boldsymbol{\theta})$ is a $k_s \times k_z$ matrix, whose (i, j)-th element denotes the density of $\boldsymbol{R}_t, \boldsymbol{F}_t$ conditional on $S_t = i, Z_t = j$ and parameter set $\boldsymbol{\theta}$. The numerators in Equation (2.7)and (2.8) denote the vector of the joint density $f(S_t, \boldsymbol{R}_t, \boldsymbol{F}_t|I_{t-1}; \boldsymbol{\theta})$ and $f(Z_t, \boldsymbol{R}_t, \boldsymbol{F}_t|I_{t-1}; \boldsymbol{\theta})$, respectively. The denominator denotes the marginal density $f(\boldsymbol{R}_t, \boldsymbol{F}_t|I_{t-1}; \boldsymbol{\theta})$ or likelihood function. The steady-state probabilities are chosen as the initial value $\xi^s_{0|0}$ and $\xi^z_{0|0}$ to start the filter and the two steps are run iteratively to obtain the likelihood function $f(\boldsymbol{R}_t, \boldsymbol{F}_t|I_{t-1}; \boldsymbol{\theta})$ and filtered probabilities sequences $\xi^s_{t|t}$ and $\xi^z_{t|t}$ for $t = 1, \ldots, T$.

We then need to draw the sequence $\tilde{\boldsymbol{S}}_T = (S_1, \ldots, S_T)$ given information at time T. Following Kim and Nelson (1999), the conditional posterior for the state variable is given by

$$g(\tilde{\boldsymbol{S}}_T|I_T;\boldsymbol{\theta}) = g(\boldsymbol{S}_T|I_T;\boldsymbol{\theta})\prod_{t=1}^{T-1} g(S_t|S_{t+1}, I_t;\boldsymbol{\theta}).$$

The sampling procedure can proceed in two steps: drawing from $g(\boldsymbol{S}_T|I_T;\boldsymbol{\theta})$ and drawing from $g(S_t|S_{t+1}, I_t;\boldsymbol{\theta})$. One can first draw S_T according to the distribution $g(\boldsymbol{S}_T|I_T;\boldsymbol{\theta})$ using $\xi^s_{T|T}$. In order to draw S_t, note that

$$\begin{aligned} g(S_t|S_{t+1}, I_t;\boldsymbol{\theta}) =& \frac{g(S_t, S_{t+1}|I_t;\boldsymbol{\theta})}{g(S_{t+1}|I_t;\boldsymbol{\theta})} \\ &\propto g(S_{t+1}|S_t;\boldsymbol{\theta})g(S_t|I_t;\boldsymbol{\theta}), \end{aligned}$$

where $g(S_{t+1}|S_t;\boldsymbol{\theta})$ is the transition probabilities of S_t and $g(S_t|I_t;\boldsymbol{\theta})$ is the filtered probabilities $\xi^s_{t|t}$. Therefore, at time t one calculates

$$g(S_t = i|S_{t+1}, I_t;\boldsymbol{\theta}) = \frac{g(S_{t+1}|S_t = i;\boldsymbol{\theta})g(S_t = i|I_t;\boldsymbol{\theta})}{\sum_{i=1}^{k_s} g(S_{t+1}|S_t = i;\boldsymbol{\theta})g(S_t = i|I_t;\boldsymbol{\theta})}$$

and compares it to a standard uniform random number. The sampling procedure is repeated for $T-1,\ldots,1$ to draw a sequence $\tilde{\boldsymbol{S}}_T$ from the conditional posterior. The same procedure applies to draw sample sequences of $\tilde{\boldsymbol{Z}}_T$.

The Dirichlet distribution is used as a conjugate prior for each column of the transition probability matrix $\boldsymbol{P}_s$. Given the sample sequence $\tilde{\boldsymbol{S}}_T$, the conditional posterior for the jth column of $\boldsymbol{P}_s$ is given by

$$g(\boldsymbol{P}_{s,j}|\tilde{\boldsymbol{S}}_T) \sim D(\theta_{j1} + \eta_{j1}, \theta_{j2} + \eta_{j2}, \ldots, \ldots, \theta_{jk_s} + \eta_{jk_s})$$

where η_{ji} refers to the number of times regime j transits to regime i counted using $\tilde{\boldsymbol{S}}_T$ and $\theta_{j1},\ldots,\theta_{jk_s}$ represents the prior parameters. The same procedure applies to $\boldsymbol{P}_z$.

Lastly, we need to draw $\{\boldsymbol{\alpha}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{k_z}, \{\boldsymbol{\beta}_i, \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\}_{i=1}^{k_s}$ given $\tilde{\boldsymbol{S}}_T, \tilde{\boldsymbol{Z}}_T, I_T$. The independent Normal inverse Wishart prior distribution is used for the regression coefficients and variance parameters. Note that both S_t and Z_t affect the inference of these parameters, one cannot derive the posterior distribution of $\{\boldsymbol{\alpha}_j\}_{j=1}^{k_z}$ and $\{\boldsymbol{\beta}_i, \boldsymbol{\mu}_i\}_{i=1}^{k_s}$ as two independent parts. Given $\{\boldsymbol{\Sigma}_j\}_{j=1}^{k_z}$, $\{\boldsymbol{\Omega}_i\}_{i=1}^{k_s}$, $\tilde{\boldsymbol{S}}_T$, $\tilde{\boldsymbol{Z}}_T$ and I_T, one can derive the posterior distribution of $\{\boldsymbol{\alpha}_j\}_{j=1}^{k_z}$ and $\{\boldsymbol{\beta}_i, \boldsymbol{\mu}_i\}_{i=1}^{k_s}$ and draw samples. Conditional on regression coefficients, state variable and the data, the conditional posterior distribution for $\{\boldsymbol{\Sigma}_j\}_{j=1}^{k_z}$, $\{\boldsymbol{\Omega}_i\}_{i=1}^{k_s}$ can also be easily obtained.

Uninformative priors are also considered and robustness checks show that the results are qualitatively invariant to the different specification of initial values and priors. Despite using different initial values, the results converge fast and become virtually the same after a burn-in period of only 1000 iterations. In this study, we use 20,000 Gibbs samples and discard the first 10,000 as burn in.

2.4 Empirical Results

2.4.1 Data

Our analysis focuses on the equities, specifically the U.S. sector portfolios which constitute the S&P 500. Since we consider an asset allocation problem, the underlying instruments should be directly investable. ETFs are ideal instruments for strategic or tactical asset allocation, and risk hedging. Due to its low expense ratio and high liquidity, they are suitable for both buy-and-hold investors and active traders. There are various S&P sector ETFs issued by different asset managers. The SPDR sector ETFs issued by State Street Global Advisors (SSGA) are used due to its relatively long trading history (started December 16, 1998), large trading volumes and total asset under management among similar products. The nine SPDR ETFs include Consumer Discretionary (XLY), Consumer Staples (XLP), Energy (XLE), Financials (XLF), Health Care (XLV), Industrials (XLI), Materials (XLB), Technology (XLK), and Utilities (XLU).[3] The weekly ETF returns data are from CRSP. For the factor portfolios, we consider the Fama and French's (1993) three factors: the market (MKT), the size (SMB) and book-to-market (HML) portfolios. The three-factor portfolios and one-month Treasury bill rate data are from Ken French's website. The excess returns are calculated according to the formula $R_{j,t} = (P_{j,t} - P_{j,t-1})/P_{j,t-1} - R_t^f$, where $P_{j,t}$ is the index value and R_t^f is the one-month Treasury bill rate. Our sample covers the period from December 16, 1998 to September 9, 2016 (924 weeks).

Table 2.1 reports summary statistics for the sector ETF and factor portfolio returns. Mean returns are all positive and lie in a range between 0.05% and 0.17% per week. Their standard deviation ranges from 1.42% to 3.93% per week. All the sector ETFs and factors portfolio returns have negative skewness and high kurtosis except that Consumer Discretionary, Financials and HML have positive skewness during the sample period. According to the Jarque-Bera test, all the sector ETFs and factor returns are strongly rejected to be normally distributed. There is some evidence of serial correlation in the MKT, SMB, and HML returns at a weekly frequency. Sector ETFs also seem to be weakly autocorrelated with the exception of Materials, Technology, and Utilities sectors. All the time series appear to have time-varying volatility according to the Engle's ARCH test. Markov regime-switching models are well known for their capability to capture heteroskedasticity, fat tails, and negative skewness in financial time series. The descriptive statistics suggest that they might be appropriate models to fit the data.

[3] S&P and MSCI have made numerous changes in the Global Industry Classification Standard (GICS) over the years. To keep up with these GICS changes, SSGA launched two new sector ETFs: Communication Services (XLC) and Real Estate (XLRE) in recent years. During the period (December 16, 1998 to September 9, 2016) we consider, these Sector ETFs have non-overlapping holdings and divide the S&P into nine index funds.

Index	Mean	SD	Min	Max	Skew	Kurt	JB	p-Val.	ARCH	p-Val.	LB-Q	p-Val.
Panel A: Summary statistics												
Consumer Discretionary	0.15	3.08	-14.74	18.31	0.09	7.19	682.47	0.00	422.60	0.00	22.48	0.01
Consumer Staples	0.09	1.99	-13.35	8.28	-0.56	7.44	816.28	0.00	783.16	0.00	24.42	0.01
Energy	0.17	3.47	-25.21	17.73	-0.54	7.45	815.64	0.00	586.98	0.00	21.94	0.02
Financials	0.08	3.93	-23.97	32.52	0.76	16.18	6809.37	0.00	813.97	0.00	43.71	0.00
Health Care	0.13	2.44	-18.60	9.49	-0.78	9.34	1650.71	0.00	1134.34	0.00	18.12	0.05
Industrials	0.13	2.90	-17.75	13.97	-0.37	6.98	635.64	0.00	560.81	0.00	21.67	0.02
Materials	0.15	3.29	-14.96	15.17	-0.21	5.53	256.05	0.00	428.43	0.00	15.71	0.11
Technology	0.08	3.39	-18.38	14.01	-0.32	5.59	276.53	0.00	458.00	0.00	10.92	0.36
Utilities	0.11	2.41	-19.82	11.82	-0.76	9.40	1676.75	0.00	1122.57	0.00	13.23	0.21
Market	0.10	2.56	-18.00	12.61	-0.56	8.38	1167.96	0.00	704.11	0.00	25.08	0.00
Small Minus Big	0.08	1.42	-10.71	7.00	-0.53	9.82	1846.43	0.00	951.48	0.00	25.11	0.01
High Minus Low	0.05	1.52	-8.68	9.92	0.66	10.56	2279.87	0.00	754.77	0.00	31.07	0.00
	XLY	XLP	XLE	XLF	XLV	XLI	XLB	XLK	XLU	MKT	SMB	HML
Panel B: Correlation												
Consumer Discretionary	1.00											
Consumer Staples	0.62	1.00										
Energy	0.54	0.45	1.00									
Financials	0.79	0.56	0.52	1.00								
Health Care	0.68	0.60	0.47	0.64	1.00							
Industrials	0.83	0.63	0.67	0.78	0.69	1.00						
Materials	0.73	0.54	0.72	0.65	0.56	0.83	1.00					
Technology	0.70	0.40	0.44	0.58	0.63	0.70	0.54	1.00				
Utilities	0.47	0.53	0.53	0.46	0.51	0.53	0.46	0.39	1.00			
Market	0.86	0.64	0.67	0.83	0.79	0.89	0.76	0.86	0.57	1.00		
Small Minus Big	0.07	-0.14	0.10	-0.06	-0.01	0.11	0.10	0.22	-0.06	0.19	1.00	
High Minus Low	0.15	0.11	0.22	0.40	-0.06	0.18	0.24	-0.24	0.13	0.04	-0.23	1.00

Table 2.1: *Summary statistics of sector ETF and Fama-French three-factor returns.* The table (Panel A) reports the summary statistics, the normality (Jarque-Bera test), the autoregressive conditional heteroscedastic (ARCH) (Engle's ARCH test with lags up to 4) and autocorrelation (Ljung–Box test with lags up to 10) test statistics and their corresponding p-values for the 9 sector ETFs and Fama-French three factor weekly returns over the period of December 16, 1998 to September 9, 2016 (924 weeks). Panel B reports the correlation matrix of these time series.

2.4.2 Regimes in the Joint Process of Factor Returns

Before we study the new regime switching multi-factor model, we first examine the regimes in the joint process of Fama French three factors as in the following model[4]:

$$\boldsymbol{F}_t = \boldsymbol{\mu}(S_t) + \boldsymbol{\Omega}^{1/2}(S_t)\boldsymbol{\epsilon}_t,$$

We estimate several specifications of the model for the returns on market, size, and

[4]Since the market, size, and value factor portfolios are all weakly autocorrelated, we also consider the presence of autocorrelation terms with up to 2 lags in an unreported result. The parsimonious MSIH(4) model is still preferred with lowest AIC, BIC, and HQC values.

value portfolios. A common Markov chain drives the means and variance-covariance matrix of factor returns. We consider linear and Markov regime switching models with the number of regimes up to 4. Table 2.2 reports the Akaike information criterion (AIC), Bayes–Schwartz information criterion (BIC), and Hannan–Quinn information criterion (HQC) which are used to perform the model selection. For each model we also show the results of the linearity test proposed by Davies (1977). The linear model is universally rejected at the 1% level. The four-state MSIH(4) model with state-dependent mean vector and variance-covariance matrix is selected according to its lowest AIC, BIC, and HQC values. This specification is consistent with the empirical finding of Guidolin and Timmermann (2007a) in which they estimate the Markov switching model using an MLE approach and select a four-state model for the joint process of monthly returns on the market, size, and value portfolios over a long period from 1927 to 2001.

Table 2.3 reports the estimates of the linear and four-state models. The first regime is a highly persistent bull state with an average duration of 92 weeks. In the first state the market risk premium is 13.54% per annum and statistically significant at the 0.01 level. The volatilities of all the three factor time series are modest and slightly smaller than its unconditional value as indicated in Table 2.2. Regime 2 is another highly persistent transient state with an average duration of 46 weeks, capturing stock prices during parts of the early 2000s (excluding the periods of dot-com crash). The market portfolio earns a negative mean return of -3.97% and the SMB and HML have significantly positive mean returns $(11.83\%, 9.92\%)$. Hence, both the size and value effects are strong in this regime. Volatility is close to its unconditional counterpart. Regime 3 is also a highly persistent transient state with an average duration of 46 weeks. The mean excess return on the market is positive at 6.18%, but not statistically significantly different from 0. While SMB and HML portfolios have mean returns that

Model (k_s)	Number of parameters	Log-likelihood	LR test for linearity	AIC	BIC	HQC
Panel A: single-state models						
MSI(1)	9	-5471.4126	NA	3.9541	3.9734	3.9611
Panel B: two-state models						
MSI(2)	14	-5312.7811	25.2936 (0.0000)	3.8469	3.8875	3.8616
MSIH(2)	20	-5033.0321	131.3358 (0.0000)	3.6328	3.6885	3.6463
Panel C: three-state models						
MSI(3)	21	-5237.0608	89.0048 (0.0000)	3.7937	3.8386	3.8099
MSIH(3)	33	-4933.2552	161.9122 (0.0000)	3.5832	3.6537	3.6086
Panel D: four-state models						
MSI(4)	30	-5268.7246	98.6318 (0.0000)	3.8230	3.8872	3.8462
MSIH(4)	48	-4862.9326	189.9056 (0.0000)	**3.5432**	**3.6459**	**3.5803**

Table 2.2: *Selection of multivariate regime switching model for Market, SMB and HML returns.* The table reports the log-likelihood, and the linearity test results, and Akaike information criterion (AIC) and Bayesian information criterion (BIC), Hannan–Quinn information criterion (HQC) for the multivariate Markov switching models: $\boldsymbol{F}_t = \boldsymbol{\mu}(S_t) + \boldsymbol{\Omega}^{1/2}(S_t)\boldsymbol{\varepsilon}_t$ with $k_s = 1, 2, 3, 4$. The sample period is from December 16, 1998 to September 9, 2016 (924 weeks). The models' acronyms are as follows: MS represents Markov Switching, I represents the presence of regime-dependent intercept, H represents the presence of regime-dependent variance-covariance matrix (heteroskedasticity).

are not significantly different from 0. Another difference between regime 2 and 3 lies in their correlation matrix. In regime 2, both the SMB and HML portfolios provide hedging to the market portfolio, while in regime 3 both the SMB and HML portfolios are positively correlated with the market. Regime 4 is a high-volatility crisis regime with an annualized volatility of 38.31% for market, 21.88% for size portfolio and, 26.25% for value portfolio, which double or triple their corresponding average value. All the market, size and value premia are negative ($-32.31\%, -3.13\%, -2.67\%$) but not precisely estimated. This state captures the periods of two recent major crisis: the dot-com bubble and 2008 financial crisis.

To further illustrate the economic interpretation of these states, Figure 2·1 plots

Panel A: Single State Model			
	Market Portfolio	SMB Portfolio	HML Portfolio
1.Mean	0.0533	0.0405*	0.0266
2.Correlations/Volatilities			
Market Portfolio	0.1848***		
SMB Portfolio	0.1893***	0.1024***	
HML Portfolio	0.0369	-0.2289***	0.1098***

Panel B: Four State Model				
	Market Portfolio	SMB Portfolio	HML Portfolio	
1.Mean				
Regime 1	0.1521***	0.0297	0.0426**	
Regime 2	-0.0397	0.1183**	0.0992*	
Regime 3	0.0618	0.0083	-0.0517	
Regime 4	-0.3231	-0.0313	-0.0267	
2.Correlations/Volatilities				
Regime 1				
Market Portfolio	0.1079***			
SMB Portfolio	0.4386***	0.0723***		
HML Portfolio	-0.0100	-0.1489***	0.0522***	
Regime 2				
Market Portfolio	0.1829***			
SMB Portfolio	-0.1592**	0.1074***		
HML Portfolio	-0.5998	-0.2011***	0.1277***	
Regime 3				
Market Portfolio	0.1902***			
SMB Portfolio	0.3901***	0.0894***		
HML Portfolio	0.2933***	0.0667	0.0912***	
Regime 4				
Market Portfolio	0.3831***			
SMB Portfolio	0.1484	0.2188***		
HML Portfolio	0.2870**	-0.4583***	0.2625***	
	Regime 1	Regime2	Regime 3	Regime 4
3.Transition matrix				
Regime1	0.9892***	0.0050	0.0201*	0.0003
Regime2	0.0001	0.9728***	0.0001	0.0670*
Regime3	0.0107*	0.0001	0.9728***	0.0214
Regime4	0.0000	0.0221*	0.0070	0.9112***

* significance at 10% level, ** significance at 5%, *** significance at 1%.

Table 2.3: *Parameter estimates of single and four-state multivariate Markov regime-switching model for Market, SMB and HML returns.* The model is $\boldsymbol{F}_t = \boldsymbol{\mu}(S_t) + \boldsymbol{\Omega}^{1/2}(S_t)\boldsymbol{\epsilon}_t$ with $k_s = 4$. The sample period is from December 16, 1998 to September 9, 2016 (924 weeks).

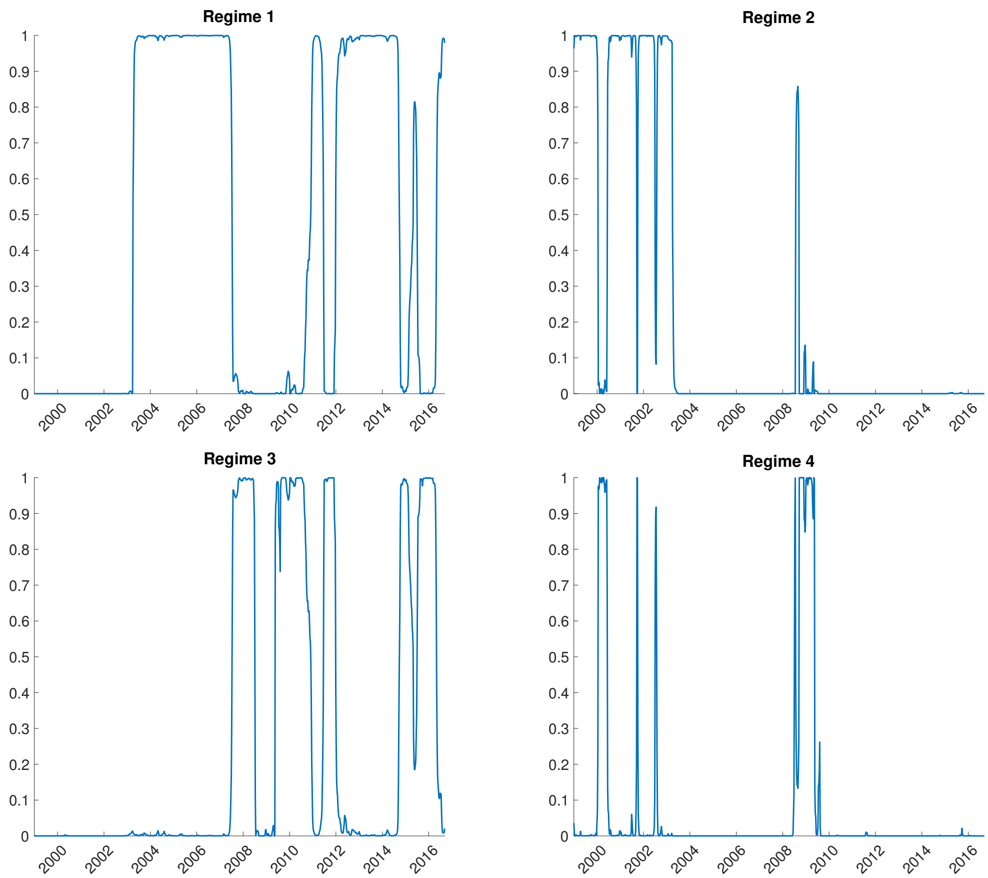

Figure 2·1: *Smoothed probabilities from four-state Markov regime switching joint model for Market Portfolio, SMB and HML returns.* The figure plots the smoothed probabilities of regime 1 (bull regime), regime 2 (transient regime), regime 3 (transient regime) and regime 4 (crisis regime) estimated from the MSIH(4) multivariate model $\boldsymbol{F}_t = \boldsymbol{\mu}(S_t) + \boldsymbol{\Omega}^{1/2}(S_t)\boldsymbol{\epsilon}_t$ comprising the weekly return series on Fama and French's (1993) three factors: MKT, SMB and HML. The sample period is from 12-31-1998 to 9-9-2016 (924 weeks).

the smoothed probability for the four-state Markov switching multivariate model. In particular, the bottom right plot in Figure 2·1 shows that the probability of Regime 4 spikes up during major crashes and declines in U.S. equity markets, such as bursting of dot-com bubble and 2008 financial crisis. The empirical findings suggest that a naive two-state (bull and bear) model might not be good enough to capture the different patterns and cycles in the dynamics of factor portfolio returns and a four-state regime switching model provides a better fit to the factor returns data.

2.4.3 Abnormal Return and Idiosyncratic Volatility of Individual Sector

Previous subsection demonstrated the presence of regimes in the market, size, and value factors. In order to investigate the dynamics of the idiosyncratic component of sector j, we characterize the sector j as

$$R_{j,t} = \alpha_j(Z_{j,t}) + \sum_{i=1}^{K} \beta_{ij}(S_t) F_{i,t}(S_t) + \sigma_j(Z_{j,t}) u_{j,t}, \quad j = 1, \ldots, N. \tag{2.9}$$

Following Billio et al. (2010), we consider a two-step approach[5]. In the first step, we identify S_t using factor returns as in subsection 2.4.2 and its estimate equals to the state with the highest smoothed probability

$$\hat{S}_t = \underset{i \in \{1,\ldots,k_s\}}{\arg\max} \Pr(S_t = i | \boldsymbol{F}_1, \ldots, \boldsymbol{F}_T), \quad t = 1, \ldots, T.$$

Conditional on this result, in the second step we can estimate the univariate regime-switching model described by Equation 2.9 and identify the latent state variable $Z_{j,t}$ and other parameters.

The estimation results of regime-dependent abnormal return and idiosyncratic

[5]Alternatively, we can jointly identify the state variables S_j, t and $Z_{j,t}$ using the algorithm described in Section 2.3. The dynamics of $Z_{j,t}$ are very similar. For each sector j, $Z_{j,t}$ can be well described by a low-volatility and high-volatility regime. $k_s = 3$ is selected for all sectors when we take into account the impact of regime-dependent factor loadings to determine the regime in the systematic part of sector returns.

volatility for different sector ETFs using the two-step approach are reported in Table 2.4. In all cases linearity is strongly rejected. We find that Markov regime switching models with two regimes are chosen according to the lowest BIC value for most sector ETFs' idiosyncratic returns except that a three-state model is slightly preferred for Utilities. The two states have a natural interpretation as a low-volatility regime ($Z_{j,t} = 1$) and a high-volatility regime ($Z_{j,t} = 2$). Both the two states are highly persistent for most sectors. The low-volatility regime is highly persistent for all the sectors with an average duration of 186.66 weeks and the high-volatility regime is also very persistent with an average duration of 66.98 weeks. For all sector ETFs the high volatility estimate σ_2 is at least twice the low volatility estimate σ_1. The volatility estimate of the nine sector ETFs in the idiosyncratic high volatility regime on average equals 18.05% per annum which is larger than two times the amount 7.30% per annum in the idiosyncratic low volatility regime. Moreover, the mean idiosyncratic returns are not statistically different from zero with exceptions that Consumer Staple and Utilities are associated with a mean return of 3.98% and 8.70% per annum respectively and Financials has a negative mean return of -3.93% in the low volatility regime.

The idiosyncratic volatilities of all sector ETFs seem to exhibit a strong degree of common time variations and very similar regime pattern. Figure 2·2 plots the smoothed probabilities of being in the high-volatility regime for each sector ETF's idiosyncratic returns. As reported in Table 2.5, the average correlation between different ETF idiosyncratic regime cycles is 68.60% and the average concordance index is 82.50%.[6] This implies that volatilities of residuals share a common regime after

[6]The concordance index is a measure of synchronization proposed by Harding and Pagan (2002, 2006):

$$I = \frac{1}{T}\left[\sum_{t=1}^{T} Z_{i,t}Z_{j,t} + \sum_{t=1}^{T}(1 - Z_{i,t})(1 - Z_{j,t})\right],$$

where $Z_{i,t}$ and $Z_{j,t}$ are two regime processes with binary states. The concordance index equals

	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care	Industrials	Materials	Technology	Utilities
α_1	-0.0006	**0.0398**	0.0223	**-0.0393**	0.0297	0.0079	0.0021	0.0039	**0.0870**
α_2	0.1002	-0.0165	0.0382	-0.0036	0.0621	0.0001	-0.0020	-0.0002	-0.1123
σ_1	**0.0571**	**0.0623**	**0.1109**	**0.0601**	**0.0648**	**0.0543**	**0.0883**	**0.0598**	**0.0996**
σ_2	**0.1875**	**0.1570**	**0.2304**	**0.1705**	**0.1378**	**0.1304**	**0.2386**	**0.1602**	**0.2117**
p_{11}	**0.9928**	**0.9982**	**0.9918**	**0.9914**	**0.9912**	**0.9941**	**0.9850**	**0.9972**	**0.9752**
p_{22}	**0.9776**	**0.9923**	**0.9883**	**0.9763**	**0.9856**	**0.9846**	**0.9543**	**0.9924**	**0.9221**

Table 2.4: *Two-state regime-switching model for individual ETF's idiosyncratic returns.* The table reports the estimation results for a two-state regime-switching model for ETF idiosyncratic returns. The estimated model is $R_{j,t} = \alpha_j(Z_{j,t}) + \sum_{i=1}^{K} \beta_{ij}(\hat{S}_t)F_{i,t} + \sigma_j(Z_{j,t})u_{j,t}$, where $F_{i,t}$ are the systematic risk factors, $\hat{S}_t$ is a four-state Markov chain identified using factor returns in the first step, $Z_{j,t}$ is also a two-state Markov chain (Regime 1 represents a low volatility idiosyncratic regime and Regime 2 represents a high volatility idiosyncratic regime), σ_1 (σ_2) is a Sector ETF's idiosyncratic volatility in low (high) idiosyncratic volatility state, p_{11} (p_{22}) represents an idiosyncratic Markov chain's transition probability of staying in its low (high) volatility state. Parameters marked in bold are statistically significant at the 95% confidence level. The sample period is from 12-31-1998 to 9-9-2016 (924 weeks).

	Consumer Discretionary	Consumer Staples	Energy	Financials	Health	Industrials	Materials Care	Technology	Utilities
Consumer Discretionary		0.89	0.81	0.87	0.89	0.97	0.87	0.91	0.84
Consumer Staples	0.90		0.87	0.76	0.77	0.79	0.74	0.74	0.67
Energy	0.62	0.67		0.77	0.83	0.89	0.88	0.78	0.83
Financials	0.76	0.72	0.58		0.85	0.87	0.80	0.87	0.76
Health	0.76	0.79	0.57	0.70		0.86	0.89	0.88	0.84
Industrials	0.92	0.94	0.65	0.79	0.77		0.85	0.80	0.80
Materials	0.74	0.71	0.54	0.75	0.65	0.80		0.84	0.76
Technology	0.73	0.84	0.51	0.57	0.79	0.76	0.59		0.69
Utilities	0.67	0.64	0.41	0.66	0.60	0.64	0.54	0.50	

Table 2.5: *Correlations and Concordance indices of the high-volatility regimes in individual sector ETF idiosyncratic returns.* The correlation parameters are reported in lower triangular and the concordance values are reported in the upper triangular. The sample period is from 12-31-1998 to 9-9-2016 (924 weeks).

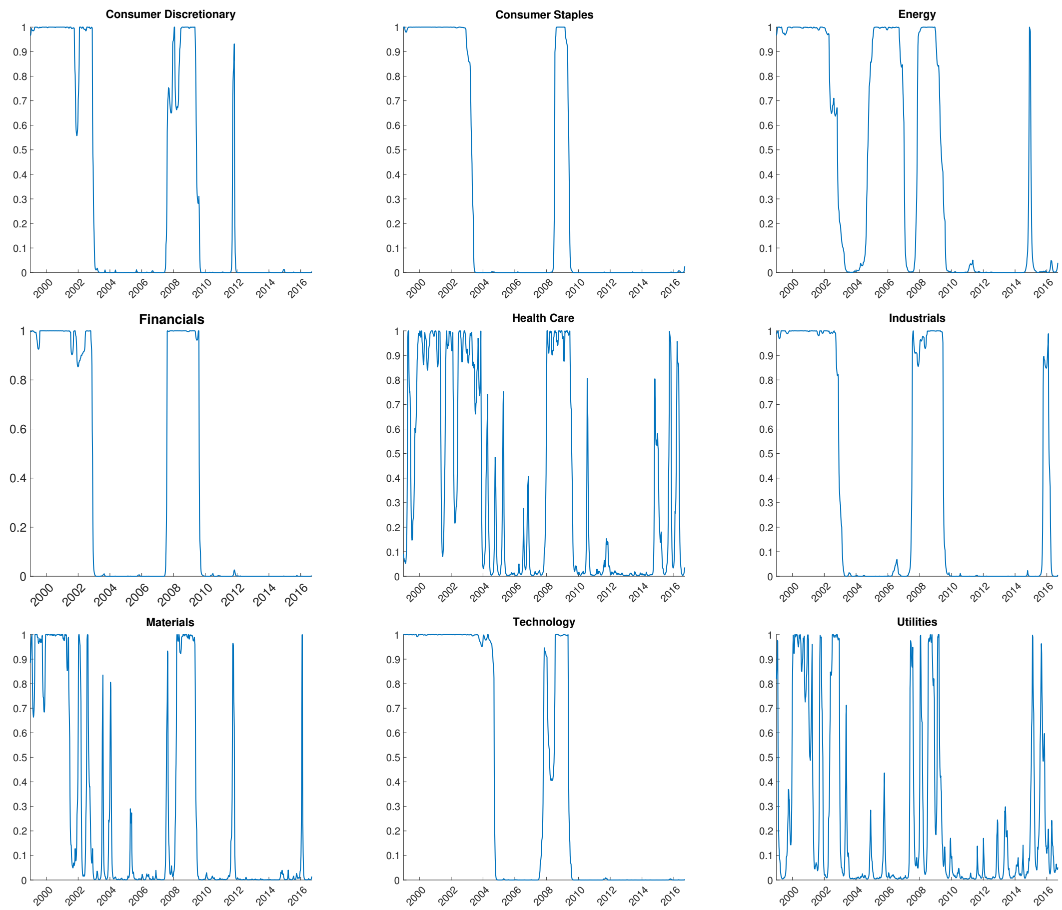

Figure 2·2: *Smoothed probabilities of being in a high-volatility state for individual sector ETFs' idiosyncratic returns under a two-state Markov regime-switching univariate model.* The estimated model is $R_{j,t} = \alpha_j(Z_{j,t}) + \sum_{i=1}^{K} \beta_{ij}(\hat{S}_t)F_{i,t} + \sigma_j(Z_{j,t})u_{j,t},$, where $F_{i,t}$ are the systematic risk factors, S_t is a four-state Markov chain identified using factor returns in the first-step, $Z_{j,t}$ is also a two-state Markov chain (Regime 1 represents a low volatility idiosyncratic regime and Regime 2 represents a high volatility idiosyncratic regime). The sample period is from 12-31-1998 to 9-9-2016 (924 weeks).

removing the common factors from returns and accounting for state-dependent factor loading. This finding is consistent with Herskovic et al. (2016), in which they find both firm-level and industry-level idiosyncratic volatilities display an extraordinary degree of co-movement. The high correlation and strong synchronization of idiosyncratic regimes indicate that it is appropriate to use a common two-state Markov chain to characterize the joint distribution of the idiosyncratic component of 9 sectors' returns.

2.4.4 A Joint Model for Sector ETF and Factor Returns

Motivated by the observations that there exist different state variables in the systematic risk factor returns and idiosyncratic component of sector ETFs, we consider the new Markov regime switching multi-factor model (2.1)-(2.2) and examine models with a combination of a different number of regimes in the systematic and idiosyncratic components. For comparison, we study the traditional regime-switching multi-factor models with a common Markov chain for both the idiosyncratic and systematic parts. BIC is used to perform model selection. We also test the linear model against the regime switching models under consideration. The saturation ratio, measured as the total number of observations divided by the total number of model parameters, is also reported.

Table 2.6 presents the outcome of this analysis. It is perhaps unsurprising that the linear model is universally strongly rejected at the significance level of 1%. The log-likelihood increases as the number of parameters increase. BIC values indicate that the new regime switching model with $k_z = 2, k_s = 3$ has the best trade-off between in-sample fit and model parsimony. The results indicate that the new regime switching model with $k_z = 2, k_s = 3$ is a competitive model to fit the joint process of the 9 sector ETFs, market, size and value portfolios.

1 when two cycles are perfectly synchronized and returns a value of 0 when they are perfectly negatively synchronized. For two independent cycles, $\hat{I}$ has an expected value close to 0.5.

Model	Number of parameters	Saturation ratio	Log-likelihood	LR test for linearity	BIC
Linear multi-factor	90	123.20	-20593	NA	3.790
Two-state multi-factor	182	60.92	-18493	4200 (0.00)	3.489
Three-state multi-factor	276	40.17	-18141	4904 (0.00)	3.504
Four-state multi-factor	372	29.81	**-17982**	5222 (0.00)	3.556
New model ($k_z = 2, k_s = 2$)	184	60.26	-18497	4192 (0.00)	3.491
New model ($k_z = 2, k_s = 3$)	224	49.50	-18263	4660 (0.00)	**3.482**
New model ($k_z = 2, k_s = 4$)	266	41.68	-18159	4868 (0.00)	3.499
New model ($k_z = 3, k_s = 3$)	282	39.32	-18095	4996 (0.00)	3.501

Table 2.6: *Selection of regime switching model for nine sector ETFs and factor returns.* The table reports the log-likelihood, and the linearity test results, and Bayesian information criterion (BIC) for the new Markov regime switching model and traditional Markov regime switching multi-factor models with a common Markov chain. k_z represents the number of regimes in the idiosyncratic returns and k_s represents the number of regimes in the systematic parts of asset returns. The sample period is from December 16, 1998 to September 9, 2016 (924 weeks).

Table 2.7 presents the estimates of the new model with $k_z = 2, k_s = 3$. As in the previous subsection, it is straightforward to interpret the two idiosyncratic regimes. The first idiosyncratic state is a moderately persistent regime with an average duration of 20.88 weeks and it is characterized by low volatility and relatively higher correlation among ETF idiosyncratic returns. The second idiosyncratic state is a less persistent high-volatility regime with an average duration of 10.95 weeks and the idiosyncratic volatility for each ETF within this state is twice or triple the magnitude of that within the low-volatility regime.

It is relatively less straightforward to interpret the three regimes for the systematic components. Regime 1 is a ‘bull’ regime characterized by large, positive market and size risk premia and low volatility for factor portfolios. Market and size factor portfolios have statistically significant mean returns of 9.57% and 4.03% per annum respectively. Value factor also has a positive mean return of 1.44%, but not statistically significant. The bull regime is a moderately persistent state and on average 19.88 weeks are spent in this state. The sector ETFs generally have smaller negative

Panel A: Idiosyncratic component	Consumer Discretion.	Consumer Staples	Energy	Financials	Health Care	Industrials	Materials	Technology	Utilities
1.Alpha									
Regime 1 (low volatility)	0.0081	**0.0309**	0.0204	**-0.0408**	0.0215	0.0009	-0.0095	-0.0147	**0.0683**
Regime 2 (high volatility)	0.0396	0.0178	0.0349	0.0085	0.0670	0.0125	0.0145	0.0424	-0.0723
2.Correlations/Volatilities									
Regime 1 (low volatility)									
Consumer Discretionary	**0.0550**								
Consumer Staples	0.0081	**0.0621**							
Energy	**-0.3227**	**-0.1702**	**0.1361**						
Financials	0.0569	**-0.1270**	**-0.4708**	**0.0599**					
Health Care	**-0.2107**	**0.1223**	**-0.1217**	-0.0425	**0.0691**				
Industrials	**0.0689**	**-0.0757**	**-0.0940**	-0.0345	**-0.1420**	**0.0534**			
Materials	**-0.1435**	**-0.1676**	**0.2933**	**-0.1606**	**-0.1966**	**0.2627**	**0.0860**		
Technology	-0.0374	**-0.2619**	**-0.2502**	**-0.1297**	**-0.3643**	**-0.1079**	**-0.1332**	**0.0608**	
Utilities	**-0.1371**	**0.2971**	**0.1083**	**-0.2305**	**0.1388**	-0.0461	**-0.0757**	**-0.2397**	**0.1095**
Regime 2 (high volatility)									
Consumer Discretionary	**0.1613**								
Consumer Staples	0.0693	**0.1487**							
Energy	**-0.2141**	-0.0455	**0.2171**						
Financials	0.0706	**-0.1021**	**-0.2313**	**0.1572**					
Health Care	0.0277	0.0738	**-0.1687**	-0.0425	**0.1451**				
Industrials	0.0710	0.0678	0.0843	**-0.1141**	-0.0567	**0.1225**			
Materials	0.0969	-0.0085	**0.3163**	**-0.1863**	-0.0618	**0.4083**	**0.1949**		
Technology	-0.0004	**-0.4462**	**-0.2106**	**-0.2786**	**-0.1871**	-0.0538	-0.0923	**0.1613**	
Utilities	**-0.1663**	0.0685	**0.1519**	**-0.1445**	-0.0240	-0.0711	**-0.1088**	**-0.1082**	**0.1731**
3.Transition matrix									
Regime1 (low volatility)	**0.9521**	**0.0913**							
Regime2 (high volatility)	**0.0479**	**0.9087**							

Panel B: Systematic component

	Consumer Discretion.	Consumer Staples	Energy	Financials	Health Care	Industrials	Materials	Technology	Utilities
1.Beta									
Regime 1 (bull)									
β_{MKT}	**1.0560**	**0.6625**	**1.0163**	**1.1293**	**0.8107**	**1.0810**	**1.1158**	**1.1232**	**0.6290**
β_{SMB}	-0.0129	**-0.2917**	0.0195	**-0.2293**	**-0.3163**	**-0.0563**	0.0647	-0.0100	**-0.2723**
β_{HML}	**-0.0999**	**-0.1852**	**0.6212**	**0.5547**	**-0.3844**	**0.1430**	**0.2681**	**-0.4197**	0.0965
Regime 2 (slow growth)									
β_{MKT}	**1.2105**	**0.4719**	**0.6700**	**1.3473**	**0.9564**	**1.0734**	**1.0499**	**0.8572**	**0.7067**
β_{SMB}	**-0.3071**	**-0.4518**	0.2137	**-0.5806**	**-0.1451**	-0.0653	-0.0410	**-0.2416**	0.0574
β_{HML}	**0.5564**	**0.2588**	**0.8312**	**0.6219**	0.1932	**0.6359**	**1.1790**	**-1.2720**	**0.7762**
Regime 3 (bear)									
β_{MKT}	**1.0552**	**0.6186**	**1.2991**	**1.1586**	**0.8776**	**1.0818**	**1.1632**	**0.9820**	**0.6989**
β_{SMB}	-0.0213	**-0.4885**	-0.2175	-0.0269	**-0.6931**	**0.2678**	**0.3122**	0.0419	**-0.6711**
β_{HML}	0.0949	**-0.1878**	**-0.3761**	**1.3399**	**-0.3475**	0.0775	-0.1173	**-0.3399**	**-0.2987**
2.Mean excess return	Market	SMB	HML						
Regime 1 (bull)	**0.0957**	**0.0403**	0.0144						
Regime 2 (slow growth)	0.0333	0.0797	0.0834						
Regime 3 (bear)	-0.1721	-0.0163	0.0409						
2.Correlations/Volatilities	Regime 1 (bull)			Regime 2 (slow growth)			Regime 3 (bear)		
	Market	SMB	HML	Market	SMB	HML	Market	SMB	HML
Market Portfolio	**0.1313**			**0.2301**			**0.3678**		
SMB Portfolio	**0.3102**	**0.0774**		0.1139	**0.1761**		0.1140	**0.1232**	
HML Portfolio	**0.1134**	-0.0312	**0.0644**	**-0.7206**	-0.4619	**0.1685**	**0.5058**	**-0.1689**	**0.2207**
3.Transition matrix	Regime 1	Regime2	Regime3						
Regime1 (bull)	**0.9497**	**0.1332**	**0.0828**						
Regime2 (slow growth)	**0.0283**	**0.7919**	**0.0616**						
Regime3 (bear)	**0.0221**	**0.0748**	**0.8556**						

Table 2.7: *Parameter estimates of the new Markov regime switching model for 9 sector ETFs and Fama-French 3 factor returns with 2 states for idiosyncratic component and 3 states for systematic component.* Parameters marked in bold are statistically significant at the 95% confidence level. The sample period is from 12-31-1998 to 9-9-2016 (924 weeks).

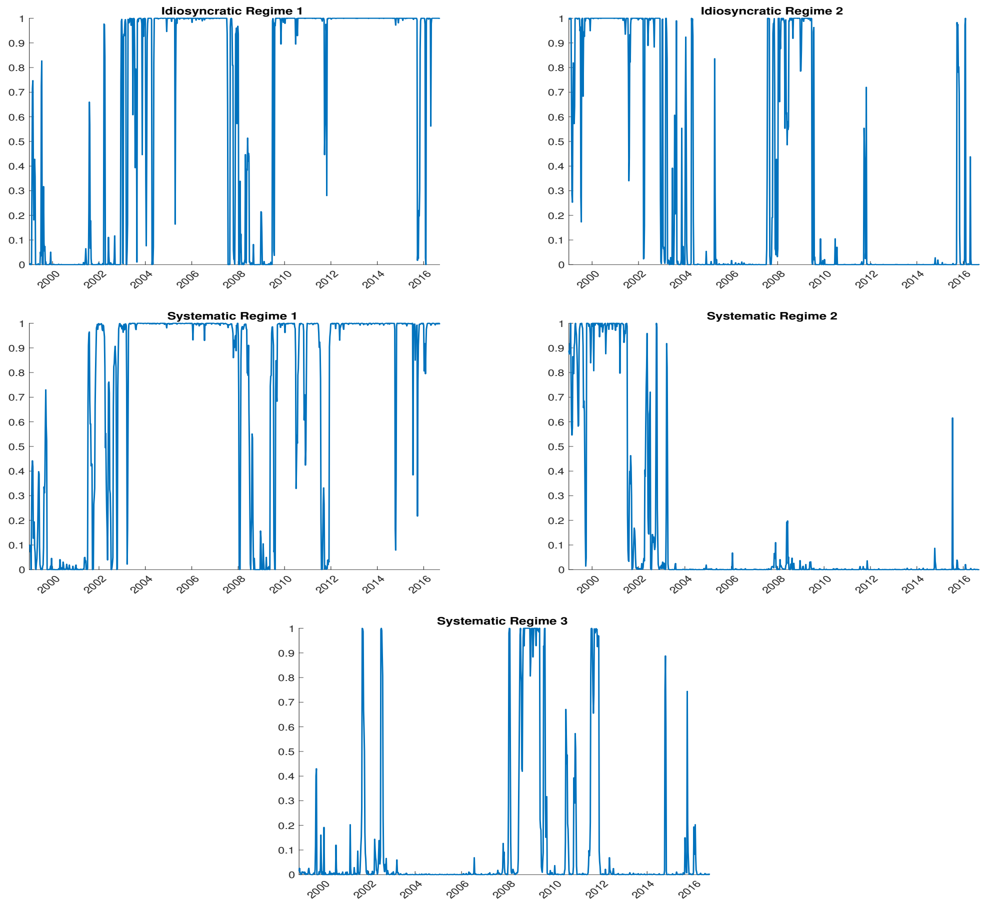

Figure 2·3: *Smoothed probabilities: the new regime switching model for 9 sector ETFs and Fama-French 3 factor returns with 2 states for the idiosyncratic component and 3 states for the systematic component.* The figure plots the smoothed probabilities of two regimes in the nine sector ETFs' idiosyncratic return and three regimes in the ETFs' systematic exposures and Fama French three factors estimated from the model

$$\boldsymbol{F}_t = \boldsymbol{\mu}(S_t) + \boldsymbol{\Omega}^{1/2}(S_t)\boldsymbol{\epsilon}_t,$$
$$\boldsymbol{R}_t = \boldsymbol{\alpha}(Z_t) + \boldsymbol{\beta}(S_t)\boldsymbol{F}_t + \boldsymbol{\Sigma}^{1/2}(Z_t)\boldsymbol{u}_t,$$

where Z_t is a two-state Markov chain characterizing the idiosyncratic components of ETFs with regime 1 corresponding to a low-volatility state and regime 2 corresponding to a high-volatility sate. S_t is a three-state Markov chain characterizing the Fama French 3 factors and sector ETFs' systematic exposure with regime 1 corresponding to a bull state, regime 2 corresponding to a slow-growth state with slow economic growth and above-average volatility, regime 3 corresponding to a bear state with high volatility. S_t is independent from Z_t. The sample period is from12-31-1998 to 9-9-2016 (924 weeks).

exposure to the size factor in this state. Regime 2 is a 'slow-growth' state characterized by above-average volatility for factor portfolios and small positive mean market excess return of 3.33%. The size and value factors have large, positive risk premia, but they are not precisely estimated. The volatilities of the factor portfolios are around twice the amount of those in the bull regime. This state has low persistence with an average duration of 4.81 weeks. The value portfolio exhibits a strong negative correlation with the market in this state. Regime 3 is a 'bear' regime characterized by high volatility and large, negative mean excess return on the market. The size and value risk premia are not significant in this state. The bear state is not very persistent with an average duration of 7 weeks.

Figure 2·3 plots the smoothed probabilities of the 2-state Z_t in the joint process of idiosyncratic ETF returns and those of the 3-state S_t that drives the factors and ETFs' systematic risk exposure. The bear regime includes the economic effects in the aftermath of the September 11 attacks, the stock market downturn of 2002, US bear market of 2007-2009, 2011 stock markets fall. The slow-growth state mainly covers the weak economic growth period of the US from 2000-2002. Both the idiosyncratic and market volatilities peak up in the state. The bull regime includes the market upturn periods not included in the other two states and takes up 74.35% of the periods considered.

2.4.5 Out-of-Sample Analysis of the Markov Regime Switching Models

Subsection 2.4.4 demonstrates that the new Markov regime switching provides a good in-sample fit to the joint process of 9 sector portfolios and Fama-French three factors. In this subsection, we conduct an out-of-sample analysis to examine the model's predictability. The first estimation window ranges from December 31, 1998 to November 23, 2007, while the forecasting period ranges from November 30, 2007 to September 23, 2011, a total of 200 weeks. An expanding window is used to estimate the regime

	New model ($k_z = 2, k_s = 3$)		Two-state Regime Switching		Three-state Regime Switching	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
Consumer Discretionary	1.6130	1.5374	1.6353	1.7043	1.6076	1.7338
Consumer Staples	1.5901	1.2199	1.6337	1.2832	1.6193	1.2842
Energy	2.5612	2.7131	2.5891	3.0018	2.5537	2.9717
Financials	1.3862	2.0422	1.4470	2.1852	1.3947	2.2164
Health Care	1.4310	1.6611	1.4526	1.6511	1.4311	1.6672
Industrials	1.2853	1.3013	1.3135	1.4074	1.2955	1.4342
Materials	2.0717	1.9656	2.1338	2.3731	2.1041	2.3617
Technology	1.7730	1.3044	1.8177	1.2556	1.7798	1.2653
Utilities	1.9189	2.1924	1.9310	2.2642	1.9102	2.2497
Average	1.7708	1.7367	1.9029	1.7726	1.9094	1.7440

Table 2.8: *Root mean square error (%) of Markov regime switching models for in-sample and out-of sample fitting.* The table reports the in-sample and out-of-sample RMSE for the new Markov regime switching model with $k_s = 3, k_z = 2$ and the conventional Markov regime switching model with $k = 2$ and $k = 3$. The in-sample period is from December 16, 1998 to November 23, 2007 and the out-of-sample period is from November 30, 2007 to September 23, 2011, on a weekly basis.

switching models and the root-mean-square error (RMSE) is used as the criterion, defined as

$$RMSE_j = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T-1}(R_{j,t+1} - \hat{R}_{j,t+1}))^2}, \qquad j = 1, \ldots, N,$$

where $\hat{R}_{j,t+1}$ represents the projected return for sector j at time $t+1$. For the traditional Markov regime switching model with a common Markov chain, $\hat{R}_{j,t+1}$ can be written as

$$\hat{R}_{j,t+1} = \sum_{i=1}^{k} \hat{P}(S_{t+1} = i | I_t; \hat{\boldsymbol{\theta}}_t)(\hat{\alpha}_{j,i} + \sum_{i=1}^{k} \hat{\boldsymbol{\beta}}_{j,i}\boldsymbol{F}_{t+1}),$$

where k represents the number of regimes for the common state variable S_t and $\hat{P}(S_{t+1} = i | I_t; \hat{\boldsymbol{\theta}}_t)$ represents the predicted probability. Similarly, for the new Markov regime switching model, $\hat{R}_{j,t+1}$ is given by

$$\hat{R}_{j,t+1} = \sum_{i_s=1}^{k_s}\sum_{i_z=1}^{k_z} \hat{P}(S_{t+1} = i_s, Z_{t+1} = i_z | I_t; \hat{\boldsymbol{\theta}}_t)(\hat{\alpha}_{j,i_z} + \sum_{i_s=1}^{k} \hat{\boldsymbol{\beta}}_{j,i_s}\boldsymbol{F}_{t+1}),$$

The analysis focuses on the performance of the new Markov regime switching model with $k_s = 3, k_z = 2$ and the conventional Markov regime switching model with $k = 2$ and 3, which perform relatively well in the in-sample fitting as in Subsection 2.4.4. Table 2.8 shows the performance of the Markov regime switching models in both the in-sample and out-of-sample fitting. The differences between the true and predicted returns under all the Markov regime switching models are small. The new Markov regime switching model generally has slightly smaller RMSE than those of the conventional Markov regime switching models. The out-of-sample RMSE under the new model is very close to its in-sample counterpart. The new model's advantage is more evident in terms of out-of-sample RMSE. This implies that the new Markov regime switching model not only provides good in-sample fit, it can also be generalizable to future data.

2.5 Asset Allocation under Regime Switching

This section studies the implications of the new regime switching multi-factor model on asset allocation. We consider an investor, who has a power utility function with a coefficient of relative risk aversion γ and investment horizon T, derives utility from his terminal wealth, W_{t+T}:[7]

$$u(W_{t+T}) = \frac{W_{t+T}^{1-\gamma}}{1-\gamma} = \frac{W_{t_B}^{1-\gamma}}{1-\gamma}. \tag{2.10}$$

The investor aims to maximize his expected utility by choosing at time t the optimal weights $\boldsymbol{\omega}_t$ among the nine sector ETFs, while $1 - \boldsymbol{\omega}_t' \iota_N$ is invested in the riskless T-bills. For simplicity, the investor is assumed to have a unit initial wealth and have no intermediate consumption.[8] His portfolio weights are rebalanced every $\varphi = \frac{T}{B}$ weeks at B equally spaced dates $t, t + \frac{T}{B}, \ldots, (B-1)\frac{T}{B}$. When $B = 1$ and

[7] For tractability consideration, the power utility function is usually a benchmark in asset allocation literature and enables comparison to previous studies.

[8] The investor's decision for saving is thus effectively exogenously specified.

$\varphi = T$, the problem collapses to a static one and the investor implements a buy-and-hold strategy.

Denote by $\boldsymbol{\omega}_{t_b} \quad (b = 0, 1, \ldots, B-1)$ the portfolio weights allocated to the risky assets and $1 - \boldsymbol{\omega}'_{t_b}\iota_N$ to T-bills at the rebalancing dates $t_b = t + b\frac{T}{B}$. The investor's asset allocation problem can be characterized by

$$\max_{\{\boldsymbol{\omega}_{t_b}\}_{b=0}^{B-1}} \quad E_t\left[\frac{W_{t_B}^{1-\gamma}}{1-\gamma}\right] \tag{2.11}$$

$$\text{s.t.} \quad W_{t_{b+1}} = W_{t_b}\exp(\varphi r^f)\{1 - \boldsymbol{\omega}'_{t_b}\iota_2 + \boldsymbol{\omega}'_{t_b}\exp(\boldsymbol{R}_{t_b,t_{b+1}})\} \tag{2.12}$$

$$\omega_{t_b}^j \geq 0 \tag{2.13}$$

where $\boldsymbol{R}_{t_b,t_{b+1}} \equiv \sum_{j=1}^{\varphi} \boldsymbol{R}_{t_b+j}$ represents the vector of excess returns on the sector ETFs, continuously compounded over the period t_b to t_{b+1}. Short-selling is not allowed.[9]

The derived utility of wealth at the rebalancing date t_b is given by

$$J(W_{t_b}, \boldsymbol{y}_{t_b}, \boldsymbol{\theta}_{t_b}, \boldsymbol{\pi}_{s,t_b}, \boldsymbol{\pi}_{z,t_b}, t_b) \equiv \max_{\{\boldsymbol{\omega}_{t_j}\}_{j=b}^{B-1}} \quad E_{t_b}\left[\frac{W_{t_B}^{1-\gamma}}{1-\gamma}\right], \tag{2.14}$$

where $\boldsymbol{\theta}_{t_b} = (\{\boldsymbol{\alpha}_{j,t_b}, \boldsymbol{\Sigma}_{j,t_b}\}_{j=1}^{k_z}, \{\boldsymbol{\mu}_{i,t_b}, \boldsymbol{\beta}_{i,t_b}\boldsymbol{\Omega}_{i,t_b}\}_{i=1}^{k_s}, \boldsymbol{P}_{s,t_b}, \boldsymbol{P}_{z,t_b})$ denotes the parameters of the Markov regime switching model, $\boldsymbol{y}_{t_b} = (\boldsymbol{R}_{t_b}, \boldsymbol{F}_{t_b})$ denotes the returns, $\boldsymbol{\pi}_{s,t_b}$ and $\boldsymbol{\pi}_{z,t_b}$ represent the vector of filtered probabilities for S_t and Z_t respectively conditional on the information up to time t_b. Under power utility, the derived utility of wealth can be simplified to (see Ingersoll, 1987 for details)

$$J(W_{t_b}, \boldsymbol{y}_{t_b}, \boldsymbol{\theta}_{t_b}, \boldsymbol{\pi}_{s,t_b}, \boldsymbol{\pi}_{z,t_b}, t_b) = \frac{W_{t_b}^{1-\gamma}}{1-\gamma} Q(\boldsymbol{y}_{t_b}, \boldsymbol{\theta}_{t_b}, \boldsymbol{\pi}_{s,t_b}, \boldsymbol{\pi}_{z,t_b}, t_b), \quad b = 0, \ldots, B-1, \gamma \neq 1.$$

Note that the underlying states are unobservable. Therefore, investors need to infer

[9] The asset allocation results allowing for short-selling are qualitatively similar.

about the underlying states and revise their beliefs according to the Bayes' rule

$$\boldsymbol{\pi}_{s,t_{b+1}}(\hat{\boldsymbol{\theta}}_{t_b}) = \frac{\left(\boldsymbol{\pi}'_{s,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{s,t_b}\right)' \boldsymbol{\pi}'_{z,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{z,t_b} \odot \boldsymbol{\eta}\left(\boldsymbol{y}_{t_{b+1}}; \hat{\boldsymbol{\theta}}_{t_b}\right)\boldsymbol{\iota}_{k_z}}{\boldsymbol{\iota}'_{k_s}\left(\boldsymbol{\pi}'_{s,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{s,t_b}\right)' \boldsymbol{\pi}'_{z,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{z,t_b} \odot \boldsymbol{\eta}\left(\boldsymbol{y}_{t_{b+1}}; \hat{\boldsymbol{\theta}}_{t_b}\right)\boldsymbol{\iota}_{k_z}}, \quad b = 0, \ldots, B-1.$$

$$\boldsymbol{\pi}_{z,t_{b+1}}(\hat{\boldsymbol{\theta}}_{t_b}) = \frac{\left(\left(\boldsymbol{\pi}'_{s,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{s,t_b}\right)' \boldsymbol{\pi}'_{z,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{z,t_b} \odot \boldsymbol{\eta}\left(\boldsymbol{y}_{t_{b+1}}; \hat{\boldsymbol{\theta}}_{t_b}\right)\right)' \boldsymbol{\iota}_{k_s}}{\boldsymbol{\iota}'_{k_s}\left(\boldsymbol{\pi}'_{s,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{s,t_b}\right)' \boldsymbol{\pi}'_{z,t_b}(\hat{\boldsymbol{\theta}}_{t_b})\hat{\boldsymbol{P}}^{\varphi}_{z,t_b} \odot \boldsymbol{\eta}\left(\boldsymbol{y}_{t_{b+1}}; \hat{\boldsymbol{\theta}}_{t_b}\right)\boldsymbol{\iota}_{k_z}}, \quad b = 0, \ldots, B-1.$$

where $\odot$ represents the element-by-element product, $\hat{\boldsymbol{\theta}}_{t_b}$ and $\hat{\boldsymbol{P}}^{\varphi}_{t_b} \equiv \Pi^{\varphi}_{i=1}\hat{\boldsymbol{P}}_{t_b}$ denotes the estimate of parameters of the regime switching model at time t_b, $\boldsymbol{\eta}\left(\boldsymbol{y}_{t_{b+1}}; \hat{\boldsymbol{\theta}}_{t_b}\right)$ is a $k_s \times k_z$ matrix whose (i,j)-th element is the density of $\boldsymbol{y}_{t_{b+1}}$ at time t_{b+1} conditional on $S_{t_{b+1}} = i$, $Z_{t_{b+1}} = j$, and the estimate of parameter set $\hat{\boldsymbol{\theta}}_{t_b}$:

$$\boldsymbol{\eta}\left(\boldsymbol{y}_{t_{b+1}}; \hat{\boldsymbol{\theta}}_{t_b}\right)_{i,j} \equiv f\left(\boldsymbol{y}_{t_{b+1}} | S_{t_{b+1}} = i, Z_{t_{b+1}} = j, \hat{\boldsymbol{\theta}}_{t_b}\right)$$
$$= (2\pi)^{-\frac{N+K}{2}} |diag(\hat{\boldsymbol{\Omega}}_i, \hat{\boldsymbol{\Sigma}}_j)^{-1}|^{\frac{1}{2}} e^{-\frac{\left(\boldsymbol{y}_{t_{b+1}} - (\hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\alpha}}_j) - (\boldsymbol{0}_K; \boldsymbol{\beta}_i \boldsymbol{F}_{t_b})\right)' diag(\hat{\boldsymbol{\Omega}}_i, \hat{\boldsymbol{\Sigma}}_j)^{-1}\left(\boldsymbol{y}_{t_{b+1}} - (\hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\alpha}}_j) - (\boldsymbol{0}_K; \boldsymbol{\beta}_i \boldsymbol{F}_{t_b})\right)}{2}}.$$

Remark. In the regime switching setup, investors are facing a time-varying investment environment driven by unobservable state variables. In addition to myopic demand, the hedging demand arises as investors update their belief about the probabilities of underlying state variables. Thus, investors' optimal portfolio choices should not only depend on the future value of the asset returns $\boldsymbol{y}_{t_b}$ but also on the future perception of the probability of being in the each of the underlying state $\boldsymbol{\pi}_{t_{b+1}}(\hat{\boldsymbol{\theta}}_{t_b})$ conditional on information up to time t_b.

2.5.1 Numerical Solutions

Previous studies have proposed a variety of approaches to solving the asset allocation problem. Barberis (2000) uses simulation methods to compute the optimal buy-and-hold and rebalancing strategies for investors with long horizons. Ang and Bekaert (2002) and Lynch (2001) employ quadrature methods to find the optimal portfolio policies for an investor. Some papers derive approximate solutions to the portfolio choice problem for infinitely-lived investors (e.g., Campbell et al., 2003; Campbell and Viceira, 1999). Wachter (2002) have derived a closed-form solution for the optimal portfolio choice problem under complete markets and mean-reverting returns.

In contrast to the observable states in Ang and Bekaert (2002), in our model investors need to infer about the underlying state and quadrature methods cannot be used to solve our problem. To solve for the optimal allocation weights accounting for investor's learning, one common approach is to approximate the integral in the expected utility functional using Monte Carlo simulation as used in Barberis (2000), Honda (2003), and Guidolin and Timmermann (2007a):

$$\max_{\boldsymbol{\omega}_t(T)} \quad \frac{1}{N_s}\sum_{i=1}^{N_s} \frac{(W_t \exp(r^f T))^{1-\gamma}}{1-\gamma} \left(1 - \boldsymbol{\omega}_t' \boldsymbol{\iota}_2 + \boldsymbol{\omega}_t' \exp\left(\boldsymbol{R}_{t,t+T}^i\right)\right)^{1-\gamma}, \tag{2.15}$$

where $\boldsymbol{R}_{t,t+T}^i$ denotes the i-th Monte Carlo simulation sample of the excess return vector from time t to $t+T$. We use $N_s = 10,000$ and a variety of investment horizon, $T = 1$ to 260 weeks.

2.6 Asset Allocation Results

In the section, we first study the buy-and-hold asset allocation strategy of an investor who solves the portfolio choice problem once at time t. Then, we introduce rebalancing and examine the out-of-sample asset allocation results. Investment horizon varies from 1 week to 5 years. Following (Guidolin and Timmermann, 2007a), the investor's coefficient of relative risk aversion γ is initially set to 5.

2.6.1 Optimal asset allocation in different regimes

As discussed in Subsection 2.4.4, the regimes in the idiosyncratic returns of sector ETFs have economic interpretations as low-volatility and high-volatility states, and

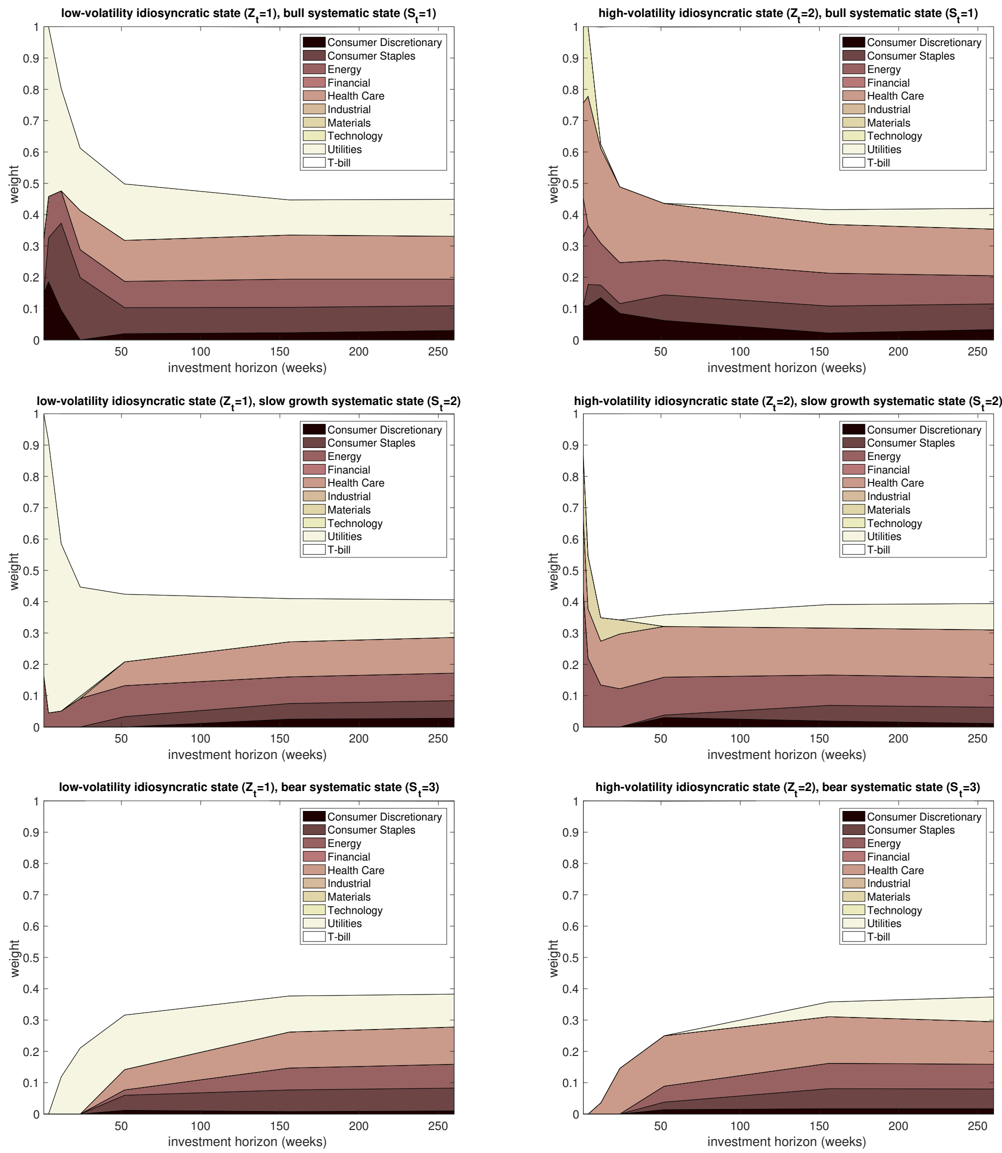

Figure 2·4: *Optimal weights as a function of the investment horizon starting from different regimes.* The graph plots a buy-and-hold investor's optimal asset allocation to the 9 sector ETFs and T-bills as a function of investment horizon within different combinations of idiosyncratic and systematic regime. The investor's relative risk aversion coefficient is set at $\gamma = 5$.

systematic regimes have economic interpretations as bull, slow-growth, and crash states. Figure 2·4 shows the optimal asset allocation for a buy-and-holder investor starting from each of the states, i.e. $(\boldsymbol{\pi}_{z,t} = \boldsymbol{e}_i, \boldsymbol{\pi}_{s,t} = \boldsymbol{e}_j)$ with $i = 1, 2$ and $j = 1, 2, 3$, but allowing for uncertainty about future regimes due to random regime switching. Given the large differences in the joint distribution of asset returns across regimes, it is perhaps unsurprising that the probabilities of the initial states have a strong impact on portfolio weights.

The bull systematic state is associated with high mean returns and low volatility for benchmark factor portfolios. The demand for sector ETFs starts at 100% at the shortest horizon and declines to around 45% at horizons for horizons greater than 1 year. Sectors including Consumer Discretionary, Consumer Staples, Energy, Health Care, Utilities are selected within the low-volatility idiosyncratic and bull systematic state. The demand for Health Care emerges at the investment horizon greater than 6 months. The idiosyncratic regime also has a significant impact on the allocation among sector ETFs. Portfolios are more diversified composing of Consumer Discretionary, Energy, Financial, Health Care and Technology at the shortest horizon within the high-volatility idiosyncratic and bull systematic state. As idiosyncratic returns start from the high-volatility state the weight on the Utilities sector becomes zero at short horizons and rises to 5% at long horizons.

In the slow growth state, the total weight on sector ETFs tends to be smaller than that under the bull systematic state and declines faster as the investment horizon grows. Energy, Health Care, and Materials are chosen at short horizons for the high-volatility idiosyncratic and slow-growth systematic state. As the idiosyncratic regime shifts to the low-volatility one, the demand for Utilities sector takes up a significant portion.

Finally, starting from the bear systematic state, 100% of the portfolio is allocated to T-bills at the shortest horizon. The proportion invested in ETFs gradually rises from 0% to 35% as the investment horizon is extended from 1 week to 5 years. The Utilities sector takes up around 10% of the portfolio with horizons longer than 3 months within the low-volatility idiosyncratic state. Its share in the optimal portfolio, to a large extent, is replaced by Health Care when the idiosyncratic returns evolve into the high-volatility regime.

2.6.2 Out-of-sample performance

A legitimate concern about the results is that although the regime switching model may produce good in-sample fits and sensible portfolio choices, the parameter estimation errors could lead to implausible and poorly determined portfolio weights when the model is used on real-time data.

To address both the concerns of model misspecification and parameter estimation error, we study how well the new regime switching model performs out-of-sample via recursive estimation and asset allocation experiment. The first estimation window is from December 31, 1998 to November 23, 2007. Using these estimates and inferred state probabilities, one can solve for the optimal portfolio weights at various investment horizons. An expanding estimation window is used to estimate the model and select the portfolio weights. We choose a sufficiently long evaluation period from November 30, 2007 to September 23, 2011, a total of 200 weeks. Note that this period covers the unprecedented 2008 financial crisis and the recovery period after the recession.

We compare the investment performance of a traditional two-state regime switching multi-factor model, a three-state regime switching multi-factor model, a new regime switching model with $k_z = 2, k_s = 3$, and a simple linear multi-factor model.[10]

[10]In order to preclude any benefits of hindsight about the optimal number of regimes, we determine the optimal number of regimes using real-time data and a flexible model with $k_z = 2, k_s = 3$ is still selected according to the lowest BIC value for the period considered.

As illustrated in Figure 2·5, regime switching lead to very different portfolio weights depending on the inferred state probabilities at a short horizon of 1 week. Relative to traditional regime switching multi-factor models, the new regime switching model tends to take more defensive positions and invests more in T-bills during market turmoil. While investing aggressively in a bull state, the new regime switching model puts a larger weight on the Energy sector than the three-state regime switching multi-factor model.

Following (Guidolin and Timmermann, 2007a) and (Tu, 2010) we use the certainty equivalent return (CER) as our performance measure. To account for the effect of the coefficient of relative risk aversion on investment performance, we consider two scenarios with risk aversion $\gamma = 5$ and $\gamma = 10$. Table 2.9 reports the investment performance for a buy-and-hold investor with various horizons, $T = 1, 4, 26, 52, 156, 260$ weeks. The results show that the new model with $k_i = 2, k_s = 3$ significantly outperformed other competing models at the short horizon with $T = 1, 4$ weeks. The certainty equivalent return of the flexible model is around 2-6% higher than that of the second-best model at these short horizons. To access its statistical significance, we do a block bootstrap experiment (50,000 trials) for the empirical distribution of CER to account for the potential serial dependence in the CER time series. The block bootstrap results show that the new regime switching model performs statistically significantly outperformed the other models over the short horizons $T = 1, 4$ at the 90% confidence level for both cases of $\gamma = 5$ and $\gamma = 10$.[11] As the investment horizon increases, the regime switching models all outperform the linear model at intermediate horizon $T = 26, 52$ weeks and the three-regime multi-factor model performs the best. But there seems no evidence that the regime switching buy-and-hold strategies outperforms the myopic strategy at long horizons $T = 156, 260$ weeks. This

[11] Two exceptions that have p-value slightly higher than 10% are the two-state model with $\gamma = 5$, $T = 1$ and the linear model with $\gamma = 10$, $T = 1$.

might be due to the fact that the regime switching model's predictability through the persistence of regimes weakens as the investment horizon increases.

2.6.3 Rebalancing

Now consider a dynamic asset allocation problem that portfolio weights can be rebalanced every $\varphi = \frac{T}{B}$ weeks at B equally spaced dates. Table 2.10 reports the investment performance for an investor allowing for weekly rebalancing ($B = T$) during the out-of-sample evaluation periods from December 7, 2008 to September 23, 2011 (200 weeks).[12] The results show that the new model with $k_z = 2, k_s = 3$ outperformed the other models in terms of CER for $\gamma = 5, 10$ and investment horizons spanning from 1 week to 5 years, with the exception of $\gamma = 10, T = 156, 260$ weeks, the three-state regime switching multi-factor model has slightly higher CER. The same procedure for block bootstrap is conducted on the CER time series to assess the statistical significance of the outperformance. The analysis of block bootstrap shows that the risk-adjusted outperformance of the new regime switching model is statistically significant at the 90% confidence interval in most cases. Both the new and traditional two-state regime switching multi-factor model statistically significantly outperformed the linear multi-factor model at intermediate to long investment horizons (T=26, 52, 156, 260) weeks, which highlights the significant economic gains for incorporating regimes in portfolio decisions.

The outperformance of the new regime switching model over the traditional two-state regime switching model might be due to the fact that the new model better fits the dynamics of asset returns. Figure 2·5 shows that the portfolio weights under the new regime switching model change abruptly as the underlying state variable probabilities change while other models do not detect the shift in the underlying state. For example, the new regime switching strategy holds a significant amount of

[12] The dynamic asset allocation with monthly rebalancing yields qualitatively similar results.

Model	Mean (%)	CER (%)	ΔCER (%)	p-value	Mean (%)	CER (%)	ΔCER (%)	p-value
	$\gamma = 5$				$\gamma = 10$			
Panel A: T = 1 week								
Linear multi-factor	-2.13	-5.62	-7.61	0.09	-0.77	-2.51	-3.01	0.17
Two-state multi-factor	4.63	-3.71	-5.70	0.14	1.69	-3.21	-3.71	0.09
Three-state multi-factor	-2.43	-7.97	-9.96	0.02	-0.39	-7.21	-7.71	0.02
New model ($k_i = 2, k_s = 3$)	**5.75**	**1.99**	NA	NA	**2.61**	**0.50**	NA	NA
Panel B: T = 4 weeks								
Linear multi-factor	-2.59	-7.28	-7.31	0.07	-1.02	-3.28	-3.48	0.10
Two-state multi-factor	0.89	-4.49	-4.52	0.10	0.03	-3.22	-3.42	0.09
Three-state multi-factor	0.54	-3.06	-3.09	0.09	0.58	-2.14	-2.34	0.10
New model ($k_i = 2, k_s = 3$)	**2.33**	**0.03**	NA	NA	**1.36**	**0.20**	NA	NA
Panel C: T = 26 weeks								
Linear multi-factor	-1.56	-7.34	-2.97	0.01	-0.54	-3.16	-1.78	0.16
Two-state multi-factor	-0.88	-5.99	-2.62	0.05	-0.23	-2.57	-1.19	0.22
Three-state multi-factor	**0.19**	**-2.33**	1.04	0.02	**0.32**	**-0.88**	0.50	0.07
New model ($k_i = 2, k_s = 3$)	-0.86	-3.37	NA	NA	-0.21	-1.38	NA	NA
Panel D: T = 52 weeks								
Linear multi-factor	-0.87	-4.51	-1.96	0.03	-0.31	-1.92	-0.94	0.15
Two-state multi-factor	-0.67	-3.33	-0.78	0.08	-0.18	-1.39	-0.41	0.19
Three-state multi-factor	**-0.15**	**-1.85**	0.70	0.01	**0.07**	**-0.69**	0.29	0.04
New model ($k_i = 2, k_s = 3$)	-0.75	-2.55	NA	NA	-0.04	-0.98	NA	NA
Panel E: T = 156 weeks								
Linear multi-factor	**3.21**	**2.76**	0.57	0.00	1.99	**1.57**	0.07	0.12
Two-state multi-factor	2.74	2.43	0.24	0.00	1.72	1.39	-0.11	0.08
Three-state multi-factor	2.47	2.21	0.02	0.38	**2.66**	1.44	-0.06	0.08
New model ($k_i = 2, k_s = 3$)	2.44	2.19	NA	NA	2.59	1.50	NA	NA
Panel E: T = 260 weeks								
Linear multi-factor	**2.40**	**1.84**	0.34	0.00	2.11	1.10	-0.32	0.00
Two-state multi-factor	2.00	1.62	0.12	0.00	3.49	1.18	-0.24	0.01
Three-state multi-factor	1.74	1.45	-0.05	0.15	**5.13**	**1.59**	0.17	0.36
New model ($k_i = 2, k_s = 3$)	1.82	1.50	NA	NA	4.45	1.42	NA	NA

Table 2.9: *Out-of-sample performance of a buy-and-hold investor under various models* The investment period includes 200 periods from December 7, 2008 to September 23, 2011. The table displays the annualized mean return and certainty equivalent return (CER) of investment performance under various models. ΔCER equals the CER of a model minus that of the new regime switching model and the p-value measures the significance level of the difference obtained through the bootstrap experiments with 50,000 trails. Performance statistics are boldfaced when these are the best.

Model	Mean (%)	CER (%)	ΔCER (%)	p-value	Mean (%)	CER (%)	ΔCER (%)	p-value
	$\gamma = 5$				$\gamma = 10$			
Panel A: T = 1 week								
Linear multi-factor	-2.13	-5.62	-7.61	0.09	-0.77	-2.51	-3.01	0.17
Two-state multi-factor	4.63	-3.71	-5.70	0.14	1.69	-3.21	-3.71	0.09
Three-state multi-factor	-2.43	-7.97	-9.96	0.02	-0.39	-7.21	-7.71	0.02
New model ($k_i = 2, k_s = 3$)	**5.75**	**1.99**	NA	NA	**2.61**	**0.50**	NA	NA
Panel B: T = 4 weeks								
Linear multi-factor	-2.05	-5.76	-7.64	0.07	-0.75	-2.58	-3.22	0.17
Two-state multi-factor	3.66	-3.29	-5.17	0.10	1.12	-3.11	-3.75	0.09
Three-state multi-factor	-2.52	-7.05	-8.93	0.01	-0.29	-5.40	-6.04	0.04
New model ($k_i = 2, k_s = 3$)	**4.95**	**1.88**	NA	NA	**2.35**	**0.64**	NA	NA
Panel C: T = 26 weeks								
Linear multi-factor	-0.59	-5.28	-8.94	0.00	-0.13	-2.41	-4.37	0.04
Two-state multi-factor	**6.74**	1.23	-2.43	0.07	**3.40**	0.15	-1.81	0.10
Three-state multi-factor	0.60	-4.63	-8.29	0.00	2.50	-1.63	-3,59	0.09
New model ($k_i = 2, k_s = 3$)	6.47	**3.66**	NA	NA	3.35	**1.96**	NA	NA
Panel D: T = 52 weeks								
Linear multi-factor	-0.00	-3.15	-8.06	0.00	0.14	-1.30	-3.92	0.00
Two-state multi-factor	**7.40**	4.01	-0.90	0.09	**4.22**	2.34	-0.28	0.28
Three-state multi-factor	1.91	-3.41	-8.32	0.00	3.91	-0.68	-3.30	0.01
New model ($k_i = 2, k_s = 3$)	7.19	**4.91**	NA	NA	3.73	**2.62**	NA	NA
Panel E: T = 156 weeks								
Linear multi-factor	3.60	2.76	-5.97	0.00	1.89	1.55	-3.08	0.00
Two-state multi-factor	10.94	8.24	-0.49	0.07	6.78	5.70	1.09	0.01
Three-state multi-factor	7.43	5.70	-3.03	0.00	**9.13**	**6.16**	1.53	0.04
New model ($k_i = 2, k_s = 3$)	**10.31**	**8.73**	NA	NA	5.13	4.63	NA	NA
Panel E: T = 260 weeks								
Linear multi-factor	3.63	2.40	-5.99	0.00	1.87	1.38	-2.99	0.00
Two-state multi-factor	10.94	8.24	-0.15	0.09	7.58	5.68	1.31	0.00
Three-state multi-factor	7.81	4.65	-3.74	0.00	**9.62**	**5.77**	1.40	0.00
New model ($k_i = 2, k_s = 3$)	**11.17**	**8.39**	NA	NA	5.31	4.37	NA	NA

Table 2.10: *Out-of-sample performance of an investor allowing for weekly rebalancing.* The investment period includes 200 periods from December 7, 2008 to September 23, 2011. The table displays the annualized mean return and certainty equivalent return (CER) of investment performance under various models. ΔCER equals the CER of a model minus that of the new regime switching model and the p-value measures the significance level of the difference obtained through the bootstrap experiments with 50,000 trails. Performance statistics are boldfaced when these are the best.

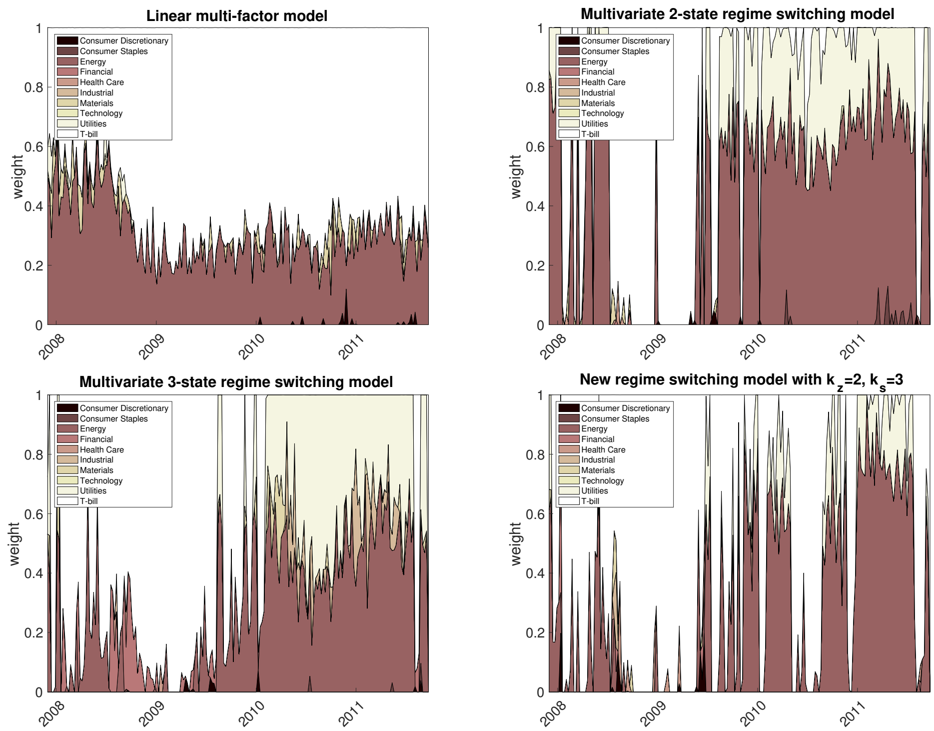

Figure 2·5: *Out-of-sample optimal weights for a buy-and-hold investor with 1-week horizon under various models.* The graph plots a buy-and-hold investor's optimal asset allocation to the 9 sector ETFs and T-bills under a linear multi-factor model, a multivariate 2-state regime switching model, a multivariate 3-state regime switching model, the new regime switching model with $k_z = 2, k_s = 3$. The evaluation period is from November 30, 2007 to September 23, 2011, a total of 200 weeks. The investor's relative risk aversion coefficient is set at $\gamma = 5$.

cash in mid 2010 while all the other models make little changes to their total risky position. As a result, large losses are avoided or alleviated during turbulent periods. The new regime switching model assumes there are three states in the systematic components of joint returns, which allows it to better capture the different patterns and magnitude in each regime. Note that the three-state regime switching multi-factor performs poorly in terms of CER at short to intermediate horizons. This implies that a naive three-state or higher-state regime switching model can easily produce over-parameterization and overfitting. Its poor out-of-sample performance provides some evidence against this model specification. The new regime switching model mitigates the issue of over-parameterization by allowing the number of idiosyncratic regimes smaller than that of the systematic risk components. The advantage of the new regime switching model over the regime switching multi-factor models decreases as the investment horizon increases. This is because the realized returns get smoothed and the severe short-period losses have less impact on the CER as the investment horizon increases.

2.7 Robustness Analysis

In this section, we consider a large set of systematic factors besides the market, size, and value factors that affect the stocks returns. We begin with the momentum factor Carhart (1997) and profitability and investment factors Fama and French (2015). The idiosyncratic volatilities under the extended multifactor models still exhibit a strong degree of co-movement.

In addition to these common factors, the liquidity and volatility factors can also be plausible candidates to explain the non-linear pattern in the idiosyncratic volatility. Here we use the spread between the US 3-month LIBOR and 3-month treasury bill as a liquidity proxy. Moreover, we consider two volatility factors used in previous studies: change in the VIX index as a new systematic factor as in Ang et al. (2006) and the common idiosyncratic volatility (CIV) factor as in Herskovic et al. (2016). To construct the CIV factor, we use all CRSP stock daily returns and compute the value-weighted monthly idiosyncratic volatility estimates with respect to the Fama and French three-factor model. Then we can get a firm's exposure to the CIV innovations as its CIV beta and sort the stocks into quintile according to their CIV-beta each year. The CIV factor is constructed as a net zero investment portfolio that long the highest quintile portfolio and short the lowest quintile portfolio.

We add these liquidity and volatility factors one by one to the Markov switching multi-factor models. However, adding these factors does not eliminate the strong regime switching pattern in the standard deviation of the residuals with respect to the extended multifactor factor model. This confirms that the strong degree of common time variation in the idiosyncratic volatilities of different sector ETFs can not be fully explained by exposure to systematic liquidity and volatility factors. This might suggest the presence of a common latent factor with regime-switching behavior across different sectors. However, identifying this common latent factor is beyond the scope of this research.[13]

2.8 Conclusion

This chapter studies regimes in the sector ETFs and market, size and value factor portfolios. The results show that regimes not only exist in systematic components of sector ETFs driven by common risk factors and factor loading but also in idiosyncratic return volatility. The idiosyncratic volatilities exhibit similar regime cycles across different sectors. This suggests a common state variable in the idiosyncratic component of the joint sector ETF returns, which can be described as a low-volatility idiosyncratic regime and a high-volatility idiosyncratic regime. While the systematic components of sector ETF returns are better captured by a three-state Markov chain, which can be interpreted as bull, bear and slow-growth states. An innovative regime switching multi-factor model is thus proposed by introducing two different state variables for the idiosyncratic and systematic parts of asset returns. Our empirical analysis shows that the new regime switching model has a better trade-off between in-sample fit and model parsimony than conventional regime switching multi-factor models with a common Markov chain. A Gibbs sampling method is also introduced to deal with the computational challenge due to the introduction of multiple Markov

[13] A related study is Akay et al. (2013), in which they uncover the common latent factor in hedge fund index returns using a dynamic Markov regime switching state-space model.

Chains and a large number of risky assets.

Furthermore, we study asset allocation implications under the regime switching multi-factor model framework. The asset allocation weights vary significantly across regimes and investment horizon. The out-of-sample asset allocation experiments confirm the economic importance of accounting for regimes in portfolio decisions. Investors change their portfolio weights considerably over time as they recursively update their beliefs about the underlying state probabilities. The out-of-sample performance results also show that the proposed regime switching model outperforms the traditional regime switching multi-factor models in terms of certainty equivalent return, especially at the short horizons. This highlights that correct specification of regime structure and number of regimes are of equal importance in asset allocation as accounting for regimes.

Chapter 3

A Rationale for the Clientele Effect in Money Management

3.1 Introduction

The money management industry is a very important segment of modern economies. According to the 2019 report[1] of Investment Company Institute (ICI), 44.8% of U.S. households delegate the management of their wealth to professional managers and the total net assets of U.S. mutual funds reached $17.7 trillion at year-end 2018.

Given the colossal size of the money management industry, studying the implications of a manager's portfolio management skill on investors' fund investment appears to be a critical task. A substantial literature finds empirical evidence that the mutual fund investors chase performance (e.g., Chevalier and Ellison, 1997; Sirri and Tufano, 1998). Berk and Green (2004) employ a model of competitive capital market and rational learning to explain the fund-performance relationship. They argue that fund managers with superior skills will manage more money but have the same fund returns as less-skilled managers because of decreasing return to scale. Furthermore, Berk and van Binsbergen (2015) assuming decreasing returns to scale use the value a mutual fund extracts from capital markets to measure the fund's skill and find the evidence of investment skill. They also find that investors appear to be able to identify the managers with superior abilities and invest more money in better funds. However,

[1] https://www.ici.org/pdf/2019_factbook.pdf

one thing that has been missing from the traditional performance evaluation and fund flow literature assuming decreasing returns to scale is the heterogeneity in investors' preferences and its effect on the segmentation of fund investors, i.e., the clientele effect. In markets where segmentation is caused by clientele effects, the fund's superior performance will not necessarily attract more money if stochastic dominance relations prevail among funds for certain groups of investors facing a mutually exclusive choice between a skilled and unskilled fund. In this case, certain risk-aversion types of investors will prefer not to invest in a skilled fund despite its superior performance. The size of a cohort of investors characterized by a possibly wealth dependent level of risk aversion therefore can limit the size of assets under management independent of the level of skill of the managers running the fund.

A growing literature shows that fund investors appear to segment the market and demonstrates the importance of clientele effects in fund performance evaluation. Blackburn et al. (2009) find that there are different investor clienteles in value and growth funds and risk aversion is an important attribute to differentiate these two groups of investors. They document that investors in value funds are more risk-averse than investors in growth funds. Moreover, Chan et al. (2002) find that growth managers have better abilities to generate alpha than value managers. Given the empirical evidence for clientele effects based on the heterogeneity in investor's risk aversion and manager's abilities, it should be investigated whether there is a rational explanation for the clientele effects in the money management industry as an alternative for behavioral interpretations such as investor sophistication (Barber et al., 2016).

This chapter studies the value of a manager's market timing skill to fund investors with heterogeneous risk preferences.[2] The clientele effect emerges as an endogenous

[2]Despite the widespread belief that mutual fund managers lack skill (e.g., Carhart, 1997; Fama and French, 2010), there is a growing number of studies that do find evidence of market timing skill

result. We model a skilled fund manager as endowed with privately informed information about the future market returns whose content is unknown to the investors.[3] Our study focuses on market timing skill for ease of presentation.[4] The anticipative information is always valuable to the manager and increases in the information precision (i.e., skill). However, when the manager and the investors exhibit different risk preferences, the private nature of this information can be costly and even adverse to the investors. We show that the investors whose risk aversion lies above a threshold value related to the manager's risk prudence would prefer the unskilled fund to the skilled one given a mutually exclusive choice between the two. Thus, there possibly are two distinct clienteles to skilled and unskilled funds dependent on the distribution of risk aversion relative to the threshold separating the two groups. Investors in skilled funds are uniformly more risk-tolerant than investors in unskilled funds.

We also analyze the impacts of commonly used portfolio management contracts on investors' fund investment. The management fee is typically a portion of the delegated wealth's value. In addition to this purely proportional fee contract, the compensation schemes including a performance-based fee that depends on the excess return of the managed portfolio relative to a benchmark are common in the money management industry. In the absence of performance fees, we find that investors whose relative risk aversion exceeds the relative prudence coefficient of the logarithmic fund managers always prefer the unskilled fund to the skilled fund irrespective of excess returns' measures like alpha and Sharpe ratios generated by the skilled fund. Conversely, investors with relative risk aversion smaller than the manager's relative

(e.g., Mamaysky et al., 2008; Elton et al., 2012; Kacperczyk et al., 2008).

[3]Our study does not consider the effort incentive problem and thus abstract from moral hazard consideration. We assume that investors can observe managers' skill level and risk preference. Koijen (2014) shows that the manager's skill and risk preference parameters can be estimated using the volatility of fund returns.

[4]In unreported results, we consider both stock selection and market timing skills in a multi-asset setting, and the clientele effect results are qualitatively similar.

prudence will choose the skilled funds. The clientele effect result still holds under the fulcrum and asymmetric performance fees. In contrast to the constant threshold under the purely proportional fee, in the presence of performance fees the relative risk aversion thresholds are affected by the skill, contract parameters, and market conditions. The comparative static analysis shows that the relative risk aversion threshold is substantially affected by the sensitivity of the contract with regard to the underperformance penalty. We find that including a fulcrum fee in the manager's compensation contract could lead to a higher value of manager's information to sufficiently risk-averse investors than that under option-like asymmetric fees. This suggests the use of fulcrum fee for investors who are much more risk-averse than the fund manager.

Our analysis proceeds in two steps. First, we derive and analyze optimal portfolio choices of a fund manager with private information. The informed manager receives a private signal about future market excess returns with noise.[5] In the presence of management fees, the investment problem is no longer maximizing the manager's utility function of terminal fund portfolio, but rather a composed utility function of fund performance. In particular, with asymmetric performance fees, the composed utility function is neither concave nor differentiable in the terminal value of fund portfolio. We employ the concavification technique pioneered by Aumann and Perles (1965), Carpenter (2000), Cuoco and Kaniel (2011) and Bichuch and Sturm (2014) to solve the manager's maximization problem. In the second step of the analysis, we study the value of the manager's information to investors and the clientele effect. Following Detemple and Rindisbacher (2013) (henceforth DR), we show that the public state price density (SPD) second-order stochastically dominates the private SPD. The clientele effect result follows from the second-order stochastic dominance

[5]The manager's skill is from his private information. Henceforth, informed is used interchangeably with skilled.

relationship and the fact that the composed utility function of investors (i.e., the utility derived from delegation to the manager as function of the SPD) may be concave or convex in the SPD depending on the investors' risk aversion is larger or smaller than a threshold value. We specialize our general results to the noisy return forecast model and conduct a comparative static analysis to understand the impacts of performance fees on investors' preference between skilled and unskilled funds in an empirically relevant setting. Finally, we discuss extensions of results to more general settings that fund managers have constant relative risk aversion (CRRA) preference and investors can invest in the market index, as an alternative to investing in the active funds.

Our study contributes to the growing literature on the clientele effect in the money management industry. Clientele effects are of great interest to research in behavioral finance. Prior studies attribute the clientele effects to irrationality, investor sophistication or other psychological tendencies (e.g., Barberis and Shleifer, 2003; Del Guercio and Reuter, 2014). Compared to behavioral attributes that may or may not determine fund clienteles, to derive the emergence of clienteles endogenously based purely on risk aversion is a viable rational alternative and as such of first order importance. In this chapter, we show that the investor clientele in a skilled fund is more risk-tolerant than the investor clientele in an unskilled fund. This result is consistent with some recent empirical findings and facts. Del Guercio and Reuter (2014) document that the mutual fund market is a segmented market catering to two distinct types of investors: direct-sold investors and broker-sold investors. Their results also suggest that direct-sold fund managers are more skilled than broker-sold fund managers because direct-sold funds have stronger incentives to hire managers with superior abilities. A recent survey by Investment Company Institute (2018) on the profile of mutual fund shareholders documents the difference between between broker-sold investors and self-directed investors. Investors who purchase mutual funds through the broker channel are less wealthy (with a median household income of $103,700 vs. $110,000) and a bit more risk-averse (with 40% vs. 44% reporting that they are willing to take above-average investment risk).

Second, our analysis is related to the delegated portfolio management literature. Existing theoretical research on delegated portfolio management focuses on two main areas. The first strand of literature studies how commonly observed compensation contracts affect manager' decisions (e.g., Grinblatt and Titman, 1989; Carpenter, 2000; Hugonnier and Kaniel, 2010). The second examines the optimal contract design problem (e.g., Admati and Pfleiderer, 1997; Li and Tiwari, 2009). We complement this literature by considering a different problem. Rather than solving for the optimal design of contracts in general, we analyze the impacts of performance fees on the value of the manager's private information to investors. Grinblatt and Titman (1989) argue that contracts should be designed with caps and have penalties for performance below the benchmark to mitigate the adverse risk incentives of managers. However, the manager is allowed to have a personal portfolio and hedge the management fees in their model. Even without the unrealistic assumption that the manager has personal accounts, our results suggest that the investors who are sufficiently more risk-averse than the manager may include a fulcrum fee component in the manager's compensation contract to realize higher value from the manager's superior information.

Finally, our study is closely related to DR. They develop a structural dynamic model of market timing and find that individuals with relative risk aversion greater than the relative risk prudence of a log manager will never prefer the skilled fund. However, their model does not consider the presence of compensation contracts between investors and fund managers. Sotes-Paladino and Zapatero (2017) find that fulcrum fees are able to help align the risk preference of investors and managers, which may distort the clientele effect results. Our study considers the presence of linear and non-linear performance fees and examine their impacts on investors' fund choice. Moreover, DR only consider log managers and investors need to make a binary choice between skilled and unskilled actively managed funds. In extensions to our basic setup, we consider settings that both investors and managers have CRRA utility and investors can choose to invest in a passive alternative. We show that the clientele effect still exists and it is affected by the fee structures, market conditions, and the manager's skill level. The comparative analysis shows that if the types of these contracts are not properly chosen, the manager's private information would be costly and even detrimental to some fund investors. Our results suggest that highly risk-averse investors should employ skilled fund managers with a linear performance-based contract in order to benefit from the manager's superior abilities.

This chapter is based on my co-authored paper Hu and Rindisbacher (2018). It is organized as follows. Section 3.2 describes the economic setup. It also presents the investment problems of managers and investors. Section 3.3 solves the portfolio optimization problem of an informed fund manager under commonly observed performance contracts. Section 3.4 analyzes the value of manager's private information and the clientele effect. Section 3.5 specializes the general results to the noisy return forecast model. Section 3.6 provides a detailed numerical analysis of the value of manager's information and investor's fund choice under fulcrum and asymmetric performance fees. Section 3.7 studies extensions of the basic model. Conclusions are in Section 3.8.

3.2 Model

This section describes the economic setup and the portfolio management problems of managers and investors. Financial markets and the information structure are described in Section 3.2.1. Section 3.2.2 introduces the private information price of risk. The agents and their risk preference are described in 3.2.3. The manager's optimization problem is described in Section 3.2.5. Section 3.2.6 describes the investor's problem.

3.2.1 Financial Markets and Information Structure

Financial markets are represented by a risky market portfolio (or stock) and a riskless bond. We work with the following model of timing information considered by DR. The instantaneous market excess return (dR^m_v) and the gross market excess return ($S^m_{\tau_{i-1},\tau_i}$) over a period $[\tau_{i-1}, \tau_i)$ are given by

$$dR^m_v = \sigma^m_v(\theta^m_v dv + dW^m_v) \quad \text{and} \quad S^m_{\tau_{i-1},\tau_i} = \exp\left(\int_{\tau_{i-1}}^{\tau_i} dR^m_v - \frac{1}{2}\int_{\tau_{i-1}}^{\tau_i} (\sigma^m_v)^2 dv\right) \quad (3.1)$$

where σ^m is positive and bounded away from zero. The volatility coefficient σ^m, the public market price of risk θ^m, and the interest rate r are stochastic process adapted to public information. The bond's price dynamics are given by $dB_v = r_v B_v dv$.

An informed agent has access to private information about future market excess returns. Her private information can be represented by the filtration

$$\mathcal{G}_{(\cdot)} = \mathcal{F}^m_{(\cdot)} \bigvee \mathcal{F}^Y_{(\cdot)} \quad (3.2)$$

where $\mathcal{F}^m_{(\cdot)}$ represents the public information generated by market excess returns dR^m, and $\mathcal{F}^Y_{(\cdot)}$ represents the filtration generated by a private signal Y. We assume that the private signal has the general anticipative form

$$Y_v \equiv \sum_{i=1}^{N} G_i \mathbb{1}_{[\tau_{i-1},\tau_i)}(v), \quad (3.3)$$

where τ_i is a sequence of deterministic dates with $\tau_0 = 0, \tau_N \leq T$, $\mathbb{1}_{[\tau_{i-1},\tau_i)}(v)$ equals 1 if $v \in [\tau_{i-1}, \tau_i)$ otherwise 0, and the private signal for the period $[\tau_{i-1}, \tau_i)$ is given by

$$G_i \equiv g\left(S^m_{\tau_{i-1},\tau_i}, \zeta_i\right) \tag{3.4}$$

for some function g and random variable ζ_i, independent of the public information $\mathcal{F}^m_T$. The independent random variable ζ_i introduces noise into the private signal G_i, thus ruling out arbitrage opportunities for the informed agent within the period $[\tau_{i-1}, \tau_i)$. The private signal works as follows. At time τ_i, the informed agent observes the signal realization $G_{i+1} \equiv g\left(S^m_{\tau_i,\tau_{i+1}}, \zeta_{i+1}\right)$ and obtains the anticipative information about the gross market excess return for the period $[\tau_i, \tau_{i+1})$. As time elapses, the informed agent learns from public information about realized market returns but the signal G_{i+1} remains valuable to her even if the time is very close to τ_{i+1}. At τ_{i+1}, a new signal realization is observed and the process repeats. In this way the informed agent maintains her information advantage against those who only have public information about market realized returns.

As in reality that active funds report their realized returns with a typically coarse schedule, we assume that informed agents only need to report their realized fund returns at τ_0 and τ_N for the period $[\tau_0, \tau_N]$. $\mathcal{F}_v = \mathcal{F}^m_v \otimes \mathcal{F}^a_{\tau_0}$ denotes the public information at time $v \in [\tau_0, \tau_N)$, where $\mathcal{F}^m_v$ is the filtration generated by realized market returns and $\mathcal{F}^a_{\tau_0}$ is the filtration generated by previously reported fund return. At time τ_N, $\mathcal{F}_{\tau_N} = \mathcal{F}^m_{\tau_N} \otimes \mathcal{F}^a_{\tau_N}$. We suppose that the investment evaluation period coincides with the time interval between the reporting dates, i.e., $T = \tau_N - \tau_0$.

3.2.2 Information premium

An agent with only public information has a premium per-unit risk denoted by the public market price of risk θ^m. The private (anticipative) information changes the price of risk by the private information price of risk (PIPR)

$$\theta_v^{\mathcal{G}} \equiv \frac{1}{\sigma_v^m} \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} E\left[\int_v^{v+\epsilon} dR_t^m \,\middle|\, \mathcal{G}_v \right] - \theta_v^m = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} E\left[\int_v^{v+\epsilon} dW_t^m \,\middle|\, \mathcal{G}_v \right] \qquad (3.5)$$

for all $v \in [0, T]$. PIPR represents the incremental price of risk, relative to θ_v^m, due to private information. The informed agent's total price of risk is thus given by $\theta_v \equiv \theta_v^m + \theta_v^{\mathcal{G}}$. When the agent has no private information, $\mathcal{G}_{(\cdot)} = \mathcal{F}_{(\cdot)}^m$ and the PIPR is null. If the agent is endowed with perfect foresight, the PIPR explodes and there exists an arbitrage opportunity. Noisy private information will lead to a PIPR with finite value and there is no arbitrage opportunity.

The market excess returns have representation $dR_v^m = \sigma_v^m((\theta_v^m + \theta_v^{\mathcal{G}})dv + dW_v^{\mathcal{G}})$ under $\mathcal{G}_{(\cdot)}$, where $dW_v^{\mathcal{G}} \equiv dW_v^m - \theta_v^{\mathcal{G}} dv$ is a Brownian motion relative to the private filtration $\mathcal{G}_{(\cdot)}$.

3.2.3 Agents

We consider an economy populated by three types of agents: investors, a skilled fund manager, and an unskilled fund manager. All agents are price-takers. The investors and the unskilled fund manager have only access to public information, while the skilled fund managers have private information about future market returns $\mathcal{G}_{(\cdot)}$. We assume that an investor at time 0 needs to make a choice between delegating the portfolio management of her wealth to either the skilled fund manager or the unskilled fund manager. No additional share purchases or redemptions are allowed during the whole investment period. The fund managers do not to have any private wealth and derive utility from the management fees received at terminal date.

We assume that both skilled and unskilled fund managers have logarithmic utility function $u^M(x) = \log(x)$ and fund investors have CRRA utility with different coefficients of relative risk aversion R. Let $\mathcal{U}$ be the class of CRRA utility functions

$$u(x) = \begin{cases} \frac{x^{1-R}}{1-R} & \text{if } R > 0, R \neq 1, \\ \log(x) & \text{if } R = 1. \end{cases}$$

Section 3.7 also examines the case of managers who have CRRA utility function with $R \neq 1$.

3.2.4 Manager's Compensation Contract

As is standard in practice, we assume that a manager is compensated at time T with a management fee F_T which depends on the end-of-period value of the fund portfolio and the end-of-period value of a benchmark portfolio. Let X_T^a represent the value assets under management (AUM) at the terminal date T. We assume that the compensation of the manager is of the general form as introduced by Cuoco and Kaniel (2011)

$$\begin{aligned} F_T &= F\left(X_T^a, X_T^b; \alpha, \beta_1, \beta_2, \delta, \pi^b\right) \\ &= \alpha X_T^a - \alpha\beta_1 X_0^a \left(\frac{X_T^a}{X_0^a} - \frac{X_T^b}{X_0^b}\right)^- + \alpha\beta_2 X_0^a \left(\frac{X_T^a}{X_0^a} - \frac{X_T^b}{X_0^b}\right)^+ \\ &= \alpha X_T^a - \alpha\beta_1 \left(X_T^a - \delta X_T^b\right)^- + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right)^+, \end{aligned}$$

where α, β_1, β_2, π^b are exogenously given parameters, $\delta = X_0^a / X_0^b$. The management fee at time T consist of three parts: a regular fee αX_T^a which is proportional to the value of the fund portfolio at time T, a performance bonus $\alpha\beta_2 \left(X_T^a - \delta X_T^b\right)^+$ which depends on the excess return of the managed fund over the benchmark, and an underperformance penalty $\alpha\beta_1 \left(X_T^a - \delta X_T^b\right)^-$. We assume that $\alpha > 0$, $\beta_2 \geq \beta_1 \geq 0$. This ensures that F is increasing and convex in the fund portfolio's end-of-period value X_T^a and decreasing in the benchmark portfolio's end-of-period value X_T^b. The benchmark portfolio process X_T^b is generated by a dynamic trading strategy π_v^b, which is known to all market participants.

The contract specification is general enough to encompass typical fee structures for different types of investment companies. When the performance bonus is symmetric to the underperformance penalty, $\beta_1 = \beta_2$, the performance fee is linear in the excess return of the actively managed fund relative to the benchmark. It is known as a "fulcrum" fee. The 1970 Amendment to the Investment Advisers Act of 1940 rules that the U.S. mutual fund's performance fees must be the "fulcrum" type. The SEC approved the use of asymmetric performance fees in contracts for investment advisers of wealthy individuals in 1985. A recent study by Ma et al. (2016) argues that, even though the advisory contracts between the asset management companies and fund investors are prohibited from having asymmetric incentive fees, the compensation incentive contracts for portfolio managers are not subject to this regulatory restriction. They document that typical compensation contracts signed by the U.S. mutual fund managers are the asymmetric, option-like type. Hedge funds are not subject to the fulcrum fee requirement, and asymmetric performance fees $\beta_1 = 0, \beta_2 > 0$ are the norm. Performance-based fees were also allowed by the Labor Department for corporate pension funds in 1986 (see Cuoco and Kaniel, 2011).

3.2.5 Manager's Problem

We consider a portfolio optimization problem of a privately informed fund manager. The problem of an uninformed manager with only public information is a special case with the PIPR $\theta^{\mathcal{G}}$ is null. Managers receive an initial endowment X_0^a from investors and choose admissible trading strategies (written $\pi_v^a \in \mathcal{G}_v$) to maximize the expected utility function of her management fee. The informed manager's problem is given by

$$\max_{\pi_v^a \in \mathcal{G}_v} E\left[u^M\left(F\left(X_T^a, X_T^b\right)\right)\Big|\,\mathcal{G}_0\right] \tag{3.6}$$

subject to

$$dX_v^a = X_v^a r_v dv + X_v^a \pi_v^a \sigma_v^m ((\theta_v^m + \theta_v^{\mathcal{G}})dv + dW_v^{\mathcal{G}}), \tag{3.7}$$

$$X_v^a \geq 0, \quad \forall v \in [0, T]. \tag{3.8}$$

Note that all the coefficients are adapted to the to the private filtration $\mathcal{G}_{(\cdot)}$, the manager's investment problem collapses to a traditional portfolio optimization problem.

3.2.6 Investor's Problem

We assume that investors cannot directly invest in the financial markets and need to employ a fund manager. Suppose that investors can observe the manager's skill level and risk preference. At time 0, an investor makes a choice between delegating his wealth to either the skilled manager or the unskilled manager based on only public information. The decision to delegate is exogenous. It captures in a reduced form the choice to abstain from direct investing because of participation constraint, transaction costs or other frictions. The composed utility function of an investor who delegates his wealth to a fund manager is $v\left(X_T^{a^*}, X_T^b\right) \equiv u\left(X_T^{a^*} - F\left(X_T^{a^*}, X_T^b\right)\right)$, where $X_T^{a^*}$ is the optimal fund value at time T chosen by the fund manager and $F\left(X_T^{a^*}, X_T^b\right)$ is the management fee paid to the manager. A fund investor maximizes the expected value of his derived utility by solving

$$\max_{\mathscr{C} \in \{s,u\}} E\left[v\left(X_T^{a,\mathscr{C}^*}, X_T^b\right)\right], \tag{3.9}$$

where X_T^{a,s^*} is the optimal terminal fund value chosen by the skilled fund manager and X_T^{a,u^*} is the optimal terminal fund value chosen by the unskilled fund manager. Extensions to the basic model consider a more general case that, alternatively to employing active managers, investors can choose a passively managed index fund.

3.3 Manager's Optimal Portfolio Policies

This section solves an informed manager's portfolio optimization problem under "fulcrum" and "asymmetric" performance fees.

The portfolio optimization problem (3.6) can be restated in the static form (see Pliska, 1986; Karatzas et al., 1987; Cox and Huang, 1989, 1991):

$$\max_{X_T^a \in \mathcal{G}_T} E\left[u^M\left(F\left(X_T^a, X_T^b\right)\right)\Big|\mathcal{G}_0\right] \tag{3.10}$$

subject to

$$E\left[\xi_T^{\mathcal{G}} X_T^a \Big| \mathcal{G}_0\right] \leq X_0^a \tag{3.11}$$

and non-negativity constraints in (3.8).

Unless $\beta_1 = \beta_2$, the fund managers' objective function $u^M\left(F\left(X_T^a, X_T^b\right)\right)$ is neither concave nor differentiable in X_T^a; it cannot be solved using the usual approach. On the other hand, the fund manager's marginal utility at zero is negative infinity, which implies that the management fee $F\left(X_T^a, X_T^b\right)$ must be strictly positive at time T. It follows that $X_T^a > \underline{X}(X_T^b)$, where $\underline{X}(X^b) = \beta_1 \delta X^b/(1+\beta_1)$. The objective function $u^M(F(\cdot, X^b))$ is piecewise concave and piecewise differentiable on the interval $[\underline{X}(X^b), \infty)$, we can follow Aumann and Perles (1965), Carpenter (2000), Cuoco and Kaniel (2011) and Bichuch and Sturm (2014) in constructing the concavification $v^M(\cdot, X^b)$ of $u^M(F(\cdot, X^b))$ (that is the smallest concave function that dominates $u^M(F(\cdot, X^b))$ for all $X^a \geq \underline{X}(X^b)$. Lemma 1 and 2 below are closely based on Lemma 1 and 2 in Cuoco and Kaniel (2011).

Lemma 1. *Suppose that $X^b > 0$, $\alpha > 0$, $\beta_2 > \beta_1 \geq 0$, there exist unique $X_1(X^b)$ and $X_2(X^b)$ with*

$$\underline{X}(X^b) < X_1(X^b) < \delta X^b < X_2(X^b)$$

that satisfy the system of equations

$$\begin{cases}\alpha(1+\beta_2)u_x^M(F(X_2(X^b),X^b) = \frac{u^M(F(X_2(X^b),X^b))-u^M(F(X_1(X^b),X^b))}{X_1(X^b)-X_2(X^b)},\\ (1+\beta_1)u_x^M(F(X_1(X^b),X^b)) = (1+\beta_2)u_x^M(F(X_2(X^b),X^b)).\end{cases}$$

In particular, if marginal utility is homogeneous of degree $-R$ $(R \neq 1)$*, letting* $\eta = \left(\frac{1+\beta_2}{1+\beta_1}\right)^{1-1/R}$*, direct computation shows that*

$$X_1(X^b) = \left(\frac{(\frac{\eta}{R}-1)\frac{\beta_1}{1+\beta_1}+\eta(1-\frac{1}{R})\frac{\beta_2}{1+\beta_2}}{\eta-1}\right)\delta X^b,$$
$$X_2(X^b) = X_1(X^b) + \frac{1}{R}\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)\delta X^b.$$

For logarithmic utility

$$X_1(X^b) = \left(log\left(\frac{1+\beta_2}{1+\beta_1}\right)\right)^{-1}\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)\delta X^b + \frac{\beta_1}{1+\beta_1}\delta X^b > \underline{X}(X^b),$$
$$X_2(X^b) = X_1(X^b) + \left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)\delta X^b.$$

Lemma 2. *Suppose that* $X^b > 0$*, let* $X_1(X^b)$ *and* $X_2(X^b)$ *be as in Lemma 1 if* $\alpha > 0$*,* $\beta_2 > \beta_1 \geq 0$ *and* $X_1(X^b) = X_2(X^b) = \delta X^b$ *if* $\alpha > 0, \beta_1 = \beta_2 \geq 0$*. The concavified objective function* $v^M(\cdot, X^b)$ *of* $u^M(F(\cdot, X^b))$ *on* $[\underline{X}(X^b), \infty)$ *is given by*

$$v^M(X^a,X^b) = \begin{cases} u^M(F(X^a,X^b)) & \text{if } X \in Y(X^b),\\ u^M(F(X_1(X^b),X^b)) & \\ +\alpha(1+\beta_2)u_x^M(F(X_2(X^b),X^b))(X^a - X_1(X^b)) & \text{otherwise,}\end{cases}$$

where $Y(X^b) = [\underline{X}(X^b), X_1(X^b)] \bigcup [X_2(X^b), \infty)$*.*

As illustrated in Figure 3·1, the concavified objective function $v^M(\cdot, X^b)$ in Lemma 2 replaces part of the original non-concave function $u^M(F(\cdot, X^b))$ with a chord between $X_1(X^b)$ and $X_2(X^b)$. The slope of the chord coincides with the slope of $u^M(F(\cdot, X^b))$ at $X_1(X^b)$ and $X_2(X^b)$, which makes the function $v^M(\cdot, X^b)$ concave. $Y(X^b)$ denotes the interval in which $v^M(\cdot, X^b)$ and $u^M(F(\cdot, X^b))$ coincide. The intuition is that be-

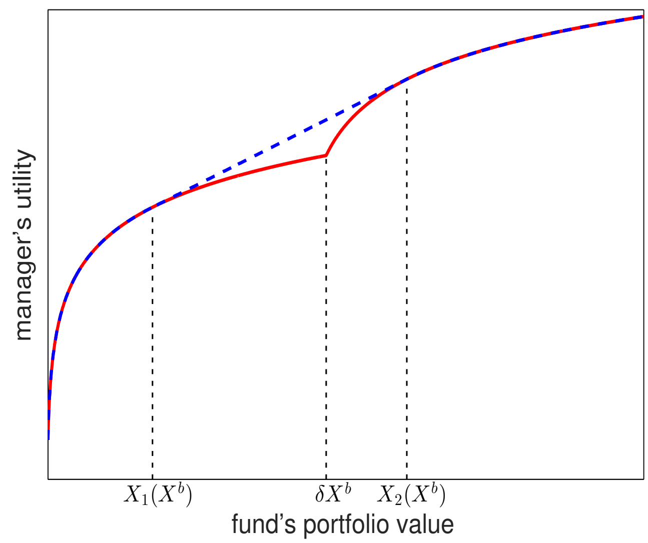

Figure 3·1: *Manager's composed utility function and the concavified function.* The figure plots the manager's composed utility function $u^M(F(\cdot, X^b))$ (red solid line) and the corresponding concavified utility function $v^M(\cdot, X^b)$ (dashed blue line) with $\alpha > 0$, $\beta_2 > \beta_1 \geq 0$.

cause the chord between $X_1(X^b)$ and $X_2(X^b)$ lies above the true objective function $u^M(F(\cdot, X^b))$, a fund portfolio's value that takes on the value $X_1(X^b)$ in some states or $X_2(X^b)$ in other states will dominate a fund portfolio's value that takes on values in the interval $(X_1(X^b), X_2(X^b))$ in some states with positive probability. Thus, the manager will never choose a fund's asset value that lies in $(X_1(X^b), X_2(X^b))$.

Since the new objective function $v^M(\cdot, X^b)$ is concave, we can use the standard method to solve the portfolio choice problem. The solution to the concavified problem also solves the original portfolio optimization problem. The solution is described formally as follows.

Proposition 1. *Suppose that the performance fee is of the fulcrum type:*

$$F(X_T^a, X_T^b) = \alpha X_T^a + \alpha\beta_2 \left(X_T^a - \delta X_T^b \right) \quad \textit{with} \quad \alpha > 0, \beta_2 \geq 0.$$

The optimal weight invested in risky asset and optimal fund value at $v \in [0, T]$ are *given by*

$$\pi_v^{a,s^*} = \frac{\theta_v^m}{\sigma_v^m} + \frac{\beta_2}{1+\beta_2}\frac{\delta X_v^b}{X_v^{a,s^*}}\left(\pi_v^b - \frac{\theta_v^m}{\sigma_v^m}\right) + \left(1 - \frac{\beta_2}{1+\beta_2}\frac{\delta X_v^b}{X_v^{a,s^*}}\right)\frac{\theta_v^{\mathcal{G}}}{\sigma_v^m}, \tag{3.12}$$

$$= \frac{\theta_v^m + \theta_v^{\mathcal{G}}}{\sigma_v^m} + \frac{\beta_2}{1+\beta_2}\frac{\delta X_v^b}{X_v^{a,s^*}}\left(\pi_v^b - \frac{\theta_v^m + \theta_v^{\mathcal{G}}}{\sigma_v^m}\right), \tag{3.13}$$

$$X_v^{a,s^*} = \frac{1}{y^{s^*}\xi_v^{\mathcal{G}}} + \frac{\beta_2 \delta X_v^b}{1+\beta_2}, \tag{3.14}$$

where $y^{s^*} = (1+\beta_2)/X_0^a$ *and* $\theta_v^{\mathcal{G}}$ *is the PIPR* (3.5)*. Correspondingly, the manager's compensation at time* T *is*

$$F(X_T^{a,s^*}, X_T^b) = \frac{\alpha X_0^a}{\xi_T^{\mathcal{G}}}. \tag{3.15}$$

Private information updates the informed manager's perceived price of risk from θ^m to $\theta^m + \theta^{\mathcal{G}}$. The first two components in (3.12) are motivated by public information and the remaining component in (3.12) is motivated by private information. If the private signal is uninformative about future market excess returns, $\theta^{\mathcal{G}} = 0$ and the optimal policy collapses to that of a manager with public information. Private signals could induce either positive or negative PIPR. When $E[dW_v^m|\mathcal{G}_v]$ is positive (negative), the PIPR is positive (negative) and the privately informed manager would invest more (less) in the market index than that of an uninformed manager. The instantaneous fund excess return generated by the informed manager is given by $d^- R_v^a = \pi_v^{a,s^*} d^- R_v^m$, where $d^- R_v^m$ represents forward integration (see Russo and Vallois, 1993). When private information induces a positive (negative) PIPR, the optimally managed fund return's volatility then increases (decreases) relative to that of a fund based only on public information. Since fund returns are only reported at the τ_0, τ_N and informed manager's trades are unobserved, this will not reveal private information to the uninformed manager.

The optimal policy can also be decomposed into two components as in Equation (3.13). The first component is the mean-variance demand and represents the manager's optimal risk taking absent the fulcrum performance fee. The second component is the benchmark-hedging demand. It could either be positive or negative, depending on whether the benchmark portfolio's weight in the risky asset π^b is higher or lower than the mean-variance demand $(\theta_v^m + \theta_v^{\mathcal{G}})/\sigma_v^m$. This component helps the manager perfectly hedge her risk exposure to the benchmark portfolio. As a result, the presence and composition of the benchmark are irrelevant to the manager's compensation at time T as given by (3.15). The manager's compensation is affected by the contract only through the proportional fee parameter α. Although the benchmark-linked incentive parameters β_2 and π^b do not affect the manager's compensation, they, together with the proportional fee parameter α, have an impact on the expected utility of fund investors' after-fee wealth. This will be analyzed in details in Section 3.4.

In line with Sotes-Paladino and Zapatero (2017), we find that the proportional fee parameter α does not affect the manager's portfolio choice. This is in contrast to the prior literature (e.g., Admati and Pfleiderer, 1997) with CARA utility function, in which the proportional fees could affect the manager's optimal risk exposure. Stronger benchmark-linked incentive β_2 leads to larger benchmark-hedging demand, which could be either a long or short position. A higher fraction of the benchmark portfolio invested in the market index increases the manager's optimal risk exposure.

Proposition 2. *Suppose that the performance fee is asymmetric:*

$$F\left(X_T^a, X_T^b\right) = \alpha X_T^a - \alpha\beta_1\left(X_T^a - \delta X_T^b\right)^- + \alpha\beta_2\left(X_T^a - \delta X_T^b\right)^+ \quad \textit{with } \alpha > 0, \beta_2 > \beta_1 \geq 0.$$

The optimal end-of-period fund value at T is given by

$$X_T^{a,s^*} = \frac{1}{y^{s^*}\xi_T^{\mathcal{G}}} + \frac{\beta_2 \delta X_T^b}{1+\beta_2}\mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} \leq \Psi(X_T^b)\}} + \frac{\beta_1 \delta X_T^b}{1+\beta_1}\mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} > \Psi(X_T^b)\}}, \tag{3.16}$$

where $\Psi\left(X_T^b\right) = \frac{\log\left(\frac{1+\beta_2}{1+\beta_1}\right)}{\delta X_T^b\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}$ *and* y^{s^*} *is a Lagrangian multiplier solving*

$$E\left[\xi_T^{\mathcal{G}} X_T^{a,s^*} \middle| \mathcal{G}_0\right] = X_0^a.$$

The fund manager's compensation at time T *is given by*

$$F(X_T^{a,s^*}, X_T^b) = \frac{\alpha(1+\beta_2)}{y^{s^*}\xi_T^{\mathcal{G}}}\mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}}\leq\Psi(X_T^b)\}} + \frac{\alpha(1+\beta_1)}{y^{s^*}\xi_T^{\mathcal{G}}}\mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}}>\Psi(X_T^b)\}}. \tag{3.17}$$

The first component in (3.16) corresponds to the optimal fund value absent performance fees ($\beta_1 = \beta_2 = 0$). The remaining two components in (3.16) are induced by the asymmetric performance fees, whose values depend on whether the normalized SPD $y^{s^*}\xi_T^{\mathcal{G}}$ is larger than $\Psi\left(X_T^b\right)$ or not. The optimal fund value at time T is thus a piecewise function of the normalized SPD $y^{s^*}\xi_T^{\mathcal{G}}$ and the end-of-period benchmark portfolio value X_T^b. Optimal end-of-period fund value X_T^{a,s^*} is greater than $X_2(X_T^b)$ and decreasing in $y^{s^*}\xi_T^{\mathcal{G}}$ until $y^{s^*}\xi_T^{\mathcal{G}}$ hits $\Psi\left(X_T^b\right)$, then X_T^{a,s^*} jumps from $1/(y^{s^*}\xi_T^{\mathcal{G}}) + \beta_2\delta X_T^b/(1+\beta_2)$ to $1/(y^{s^*}\xi_T^{\mathcal{G}}) + \beta_1\delta X_T^b/(1+\beta_1)$.

When the normalized SPD $y^{s^*}\xi_T^{\mathcal{G}}$ is smaller than $\Psi\left(X_T^b\right)$, the managed portfolio outperforms the given benchmark portfolio and the manager receives a relatively high compensation $\alpha(1+\beta_2)/(y^{s^*}\xi_T^{\mathcal{G}})$. Conversely, when the normalized SPD $y^{s^*}\xi_T^{\mathcal{G}}$ is larger than $\Psi\left(X_T^b\right)$, the managed portfolio underperforms the benchmark and the manager's compensation is then $\alpha(1+\beta_1)/(y^{s^*}\xi_T^{\mathcal{G}})$. In contrast to the fulcrum type fee case, the manager cannot completely hedge the risk induced by the asymmetric performance fees by moving up or down the risky asset in the portfolio. All the non-linear contract parameters will affect the manager's compensation as well as the derived utility of investors' after-fee wealth. As in the fulcrum type fee case, the proportional fee parameter α has no impact on the optimal fund value X_T^{a,s^*} as given by (3.16). Thus, it does not affect the optimal portfolio weight in the risky asset either.

3.4 Value of Information and the Clientele Effect

This section analyzes the incremental value of a manager's private information relative to public information and the clientele effect, which emerges as a result of investors' choices. Section 3.4.1 examines the case of fulcrum type contracts. Section 3.4.2 studies the asymmetric performance fees case.

3.4.1 Fulcrum performance contracts

We start our analysis of an uninformed investor's fund choice in the presence of symmetric fees, $F(X_T^a, X_T^b) = \alpha X_T^a + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right)$ with $\alpha > 0$ and $\beta_2 \geq 0$. The performance-related component of the management fee $F(X_T^a, X_T^b)$ is linear in the excess return of the managed fund over a benchmark. These types of contracts are known as fulcrum performance contracts and commonly observed in practice. In 1970, the amendment to the Investment Advisers Act of 1940 rules that the U.S. mutual fund performance fees must be the fulcrum type.

The (ex ante) value of private information for the fund manager can be computed as the certainty equivalent return (CER) achieved with private information in excess of the certainty equivalent return without the information advantage. It is described next.

Proposition 3. *Let process $p_{(\cdot)}^G \equiv \{p_v^G(z) : v \in [\tau_{i-1}, \tau_i)\}$ be the conditional density process of the signal G_i given public information. In the presence of fulcrum performance fees $F(X_T^a, X_T^b) = \alpha X_T^a + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right)$ with $\alpha > 0$ and $\beta_2 \geq 0$, the (ex ante) incremental value of the private signal $Y_v = \sum_{i=1}^N G_i \mathbb{1}_{[\tau_{i-1}, \tau_i)}(v)$ for the manager is*

$$V^{M,f} \equiv CER^{M,s} - CER^{M,u} = \frac{1}{2}\int_0^T E\left[\left(\theta_v^G\right)^2\right] dv = \sum_{i=1}^N E\left[\mathcal{D}_{KL}\left(p_{\tau_i}^G(G_i) \,\middle|\, p_{\tau_{i-1}}^G(G_i)\right)\right] \tag{3.18}$$

where $\mathcal{D}_{KL}\left(p_{\tau_i}^G(G_i) \,\middle|\, p_{\tau_{i-1}}^G(G_i)\right) \equiv E\left[\log \frac{p_{\tau_i}^G(G_i)}{p_{\tau_{i-1}}^G(G_i)} \,\middle|\, \mathcal{F}_{\tau_i}\right]$ is the relative entropy of the *signal. The private signal has no value to the manager if and only if the PIPR is null.*

The value of information to the fund manager (3.18) is non-negative and increasing in the absolute value of the PIPR. Proposition 3 also shows that the source of private information value is the relative entropy between an uninformed individual's beliefs about the signal at time τ_{i-1} and τ_i. The relative entropy $\mathcal{D}_{KL}\left(p^G_{\tau_i}(G_i)\,\middle|\,p^G_{\tau_{i-1}}(G_i)\right)$ measures the information gained when one updates her beliefs from the prior probability distribution $p^G_{\tau_{i-1}}(G_i)$ to the posterior probability distribution $p^G_{\tau_i}(G_i)$. If the public information at time τ_i provides valuable information about the signal G_i relative to the public information at time τ_{i-1}, the relative entropy will be positive. If the signal is unrelated to the public information, the prior and posterior probability distributions will be the same, leading the relative entropy to be zero.

Neither the fulcrum performance fees nor the proportional fees affect the value of information to the fund manager. Since the manager is able to undo any benchmark-linked incentive implemented through linear contracts, the manager's excess CER is unaffected by the power of incentives β_2 or the benchmark composition π_b. The proportional fee parameter α does not affect the manager's excess CER either. This is because the skilled (respectively unskilled) manager's compensation at time T is $\alpha X^a_0/\xi^{\mathcal{G}}_T$ (respectively $\alpha X^a_0/\xi^m_T$), α vanishes when the excess CER is computed.

Let $I(y,b) = \frac{1-\alpha-\alpha\beta_2}{y} + \frac{\beta_2\delta b}{1+\beta_2}$. The derived utility function of a fund investor is

$$u\left(X^{a^*}_T - F\left(X^{a^*}_T, X^b_T\right)\right) = u\left(I(y^*\xi_T, X^b_T)\right) \equiv v^f\left(\xi_T, X^b_T\right),$$

where $y^* = (1+\beta_2)/X^a_0$. The associated value function is $U \equiv E\left[v^f\left(\xi_T, X^b_T\right)\right]$. An unskilled (respectively skilled) fund manager optimizes her portfolio based on the public SPD ξ^m (respectively private SPD $\xi^{\mathcal{G}}$). Investors would prefer a skilled fund instead of an unskilled fund when $U^s > U^u$. Let $\mathcal{U}^s \subset \mathcal{U}$ be the subset of investors' utilities that find it better off investing in the skilled fund rather than the unskilled one

$$\mathcal{U}^s = \left\{ u \in \mathcal{U} : \Delta \equiv U^s - U^u = E\left[v^f\left(\xi_T^{\mathcal{G}}, X_T^b \right) - v^f\left(\xi_T^m, X_T^b \right) \right] > 0 \right\}.$$

Let $F^{\xi^m,b}$ (respectively $F^{\xi^{\mathcal{G}},b}$) be the cumulative distribution function (CDF) of ξ_T^m (respectively $\xi_T^{\mathcal{G}}$) conditional on the σ-algebra generated by the benchmark portfolio at time T, $\sigma(X_T^b)$. As $E\left[\xi_T^{\mathcal{G}} \middle| \sigma(X_T^b) \right] = E\left[\xi_T^m \middle| \sigma(X_T^b) \right]$ (see the proof of Proposition 4), one does not dominate the other in the sense of first-order stochastic dominance. Let $T^{\xi,b}$ be the cumulative spread between the distributions of private and public SPDs conditional on $\sigma(X_T^b)$.

Proposition 4. *The public state price density second-order stochastically dominates (SSD) the private state price density*

$$\xi_T^m \ \mathcal{SSD} \ \xi_T^{\mathcal{G}} \quad \text{and} \quad T^{\xi,b}(z) \equiv \int_{-\infty}^{z} \left(F^{\xi^{\mathcal{G}},b}(y) - F^{\xi^m,b}(y) \right) dy \geq 0 \quad \text{for all} \quad z \in \mathbb{R}^+. \tag{3.19}$$

Proposition 4 generalizes the stochastic dominance result of DR, allowing the distribution of the SPDs to be conditional on the benchmark and signal realizations for the whole investment period. The second-order stochastic dominance result follows from the fact that the ratio of private and public SPDs corresponds to the product of reciprocal of the signal's density process, $\xi_T^{\mathcal{G}} = \xi_T^m \prod_{i=1}^{N} p_{\tau_{i-1}}^{\mathcal{G}}(G_i)/p_{\tau_i}^{\mathcal{G}}(G_i)$. As ξ_T^m and $\xi_T^{\mathcal{G}}$ have the same mean conditional on $\sigma(X_T^b)$, the private SPD $\xi_T^{\mathcal{G}}$ is a mean-preserving spread of the public SPD ξ_T^m.

The SSD result has the potential to formulate a second-order stochastic dominance test to evaluate whether actively managed funds have timing skills or not controlling the effects of management fees. For example, with a linear performance-based fee, the test can be implemented using a Kolmogorov-Smirnov statistic or other statistics to quantify the spread between the empirical distribution of $\left(X_T^{a,s^*} - \frac{\beta_2}{1+\beta_2} X_T^b \right)^{-1}$ and the known parametric distribution of $\left(X_T^{a,s^*} - \frac{\beta_2}{1+\beta_2} X_T^b\right)^{-1}$.

Proposition 5 describes the value of private information to investors and the clientele effect by making use of the SSD result.

Proposition 5. *In the presence of fulcrum performance fees* $F(X_T^a, X_T^b) = \alpha X_T^a + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right)$ *with* $\alpha > 0$ *and* $\beta_2 \geq 0$, *the value of the manager's private signal* G_i (3.4) *to an investor with relative risk aversion* R *is*

$$V^f \equiv CER^s - CER^u = \frac{1}{1-R} \log \left(\frac{E\left[v^f\left(\xi_T^{\mathcal{G}}, X_T^b\right)\right]}{E\left[v^f\left(\xi_T^m, X_T^b\right)\right]} \right). \tag{3.20}$$

Let $\Delta \equiv U^s - U^u = E\left[\int_0^\infty \frac{\partial^2 v^f}{\partial z^2}(z, X_T^b) T^{\xi,b}(z) dz\right]$. *The set of investors who prefer the skilled fund* $\mathcal{U}^s = \{u \in \mathcal{U} : \Delta > 0\} = \left\{u \in \mathcal{U} : V^f > 0\right\}$ *is given by*

$$\mathcal{U}^s = \left\{ u \in \mathcal{U} : E\left[\int_0^\infty \frac{\left(\frac{1-\alpha(1+\beta_2)}{y^* z^2}\right)^2 \left[2\left(1 + \frac{\beta_2 \delta X_T^b}{1+\beta_2} \frac{y^* z}{1-\alpha(1+\beta_2)}\right) - R\right] T^{\xi,b}(z)}{\left(\frac{1-\alpha(1+\beta_2)}{y^* z} + \frac{\beta_2 \delta X_T^b}{1+\beta_2}\right)^{R+1}} dz \right] > 0 \right\}.$$

where $y^* = (1+\beta_2)/X_0^a$ *is the Lagrange multiplier,* R *is the relative risk aversion (RRA) of investors. The unskilled fund is preferred by investors with utility function in the set* $\mathcal{U}^u = \{u \in \mathcal{U} : \Delta < 0\} = \left\{u \in \mathcal{U} : V^f < 0\right\}$. *For any level of skill, there exists a value* $R^* > 0$ *and investors with coefficient of RRA exceeding* R^* *will choose the unskilled fund.*

In particular, with purely proportional fees $\beta_2 = 0$, *skilled fund returns are preferred by all investors with* $R < 2$. *Conversely, investors with relative risk aversion* $R \geq 2$ *will be better off choosing the unskilled fund irrespective of the skill level. The RRA threshold value 2 here corresponds to the relative risk prudence (RRP) of the manager with log utility.*

Expression (3.20) describes the (incremental) value of the manager's private signal to an investor. It is the analog of (3.18), except that all the contract parameters affect the investor's excess CER. The risk aversion misalignment between the manager and investors leads to a loss of the value of the manager's information that investors can exploit. Proposition 5 shows that when the misalignment is large enough, the value of the manager's private information to an investor might even be negative.

Proposition 5 also provides a characterization of potential investors in a skilled fund. It shows that investors with RRA less than the manager's RRP always prefer the skilled funds to unskilled ones. However, when investors' relative risk aversion is sufficiently high, they would choose the uninformed funds rather than the ones which have access to anticipative information. In particular, when the management fee is purely proportional to the end-of-period AUM ($\beta_2 = 0$), the manager's optimal portfolio collapses to the mean-variance demand. This leads to a fixed relative risk aversion threshold, equaling the manager's RRP, that divides the investors into different clienteles to skilled and unskilled funds. Investors with relative risk aversion less than the manager's RRP prefer the skilled fund return; conversely, investors with RRA exceeding the manager's RRP never prefer the skilled return independently of all the parameters. The result incorporates the special case of no management fee considered in DR.

In the presence of fulcrum performance fees, for any given skill level there always exist investors whose RRA exceeds a threshold value R^* will prefer the unskilled fund manager to the skilled one. R^* is determined by the equation $\Delta(R) = 0$. If the solution is not unique, we have an odd number of roots and there still exists a group of highly risk-averse investors whose RRA larger than the largest root will choose the unskilled fund. The threshold value R^* is no longer constant but depends on the manager's skill, contract parameters, and market conditions. The clientele effect is a direct result of the SSD relationship between public and private SPDs and the composed utility function of the investor $u \circ I$ is strictly concave (respectively convex) in the SPD ξ if and only if $R \geq R^*$ (respectively $R < R^*$).

3.4.2 Asymmetric performance contract

Although the Investment Company Amendments Act of 1970 places restrictions that U.S. mutual funds' advisory contracts must be the fulcrum type, many U.S. hedge funds, institutional funds, and mutual funds outside the United States employ the asymmetric fees (Golec and Starks, 2004). Furthermore, a recent study by Ma et al. (2016) document that most of U.S. mutual fund managers are offered the option-like, performance-based compensation contracts.

We assume that both the skilled and unskilled fund managers receive asymmetric performance fees $F(X_T^a, X_T^b) = \alpha X_T^a - \alpha\beta_1 \left(X_T^a - \delta X_T^b\right)^- + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right)^+$ with the same parameters $\alpha > 0, \beta_2 > \beta_1 > 0$. It is shown in (3.17) that the optimal fund value and manager's compensation are both piecewise functions of the normalized state price density and benchmark portfolio's value. Let

$$
\begin{aligned}
g(y,b) &= \frac{1}{y} + \frac{\beta_2 \delta b}{1+\beta_2} \mathbb{1}_{\{y \le \Psi(b)\}} + \frac{\beta_1 \delta b}{1+\beta_1} \mathbb{1}_{\{y > \Psi(b)\}}, \\
g^c(y,b) &= \frac{\alpha(1+\beta_2)}{y} \mathbb{1}_{\{y \le \Psi(b)\}} + \frac{\alpha(1+\beta_1)}{y} \mathbb{1}_{\{y > \Psi(b)\}},
\end{aligned}
$$

where $\Psi(b) = \frac{\log\left(\frac{1+\beta_2}{1+\beta_1}\right)}{\delta b \left(\frac{\beta_2}{1+\beta_2} - \frac{\beta_1}{1+\beta_1}\right)}$. The skilled and unskilled fund managers' compensation at time T are given by

$$
\begin{aligned}
F(X_T^{a,s^*}, X_T^b) &= g^c(y^{s^*} \xi_T^{\mathcal{G}}, X_T^b), \\
F(X_T^{a,u^*}, X_T^b) &= g^c(y^{u^*} \xi_T^m, X_T^b),
\end{aligned}
$$

where the Lagrange multipliers y^{s^*} and y^{u^*} are determined by

$$
E\left[\xi_T^{\mathcal{G}} g\left(y^{s^*} \xi_T^{\mathcal{G}}, X_T^b\right) \,\middle|\, \mathcal{G}_0\right] = X_0^a, \tag{3.21}
$$

$$
E\left[\xi_T^m g\left(y^{u^*} \xi_T^m, X_T^b\right)\right] = X_0^a. \tag{3.22}
$$

The associated utility function of a fund investor is given by

$$v^a\left(y^*\xi_T, X_T^b\right) \equiv u\left(g\left(y^*\xi_T, X_T^b\right) - g^c\left(y^*\xi_T, X_T^b\right)\right) \tag{3.23}$$

Note that y^{s^*} is a random variable that depends on the realization of the private signal G_1 and solves the Equation (3.21). By contrast, in the case of the fulcrum type fee, y^{s^*} is independent of the signal and it equals y^{u^*}.

Proposition 6. *In the presence of asymmetric performance fees:*

$$F(X_T^a, X_T^b) = \alpha X_T^a - \alpha\beta_1\left(X_T^a - \delta X_T^b\right)^- + \alpha\beta_2\left(X_T^a - \delta X_T^b\right)^+ \text{ with } \alpha > 0, \beta_2 > \beta_1 > 0,$$

the ex ante value of the private signals G_i with $i = 1, \ldots, N$ to a fund manager is

$$V^{M,a} \equiv CER^{M,s} - CER^{M,u} = E\left[\log\frac{g^c(y^{s^*}\xi_T^{\mathcal{G}}, X_T^b)}{g^c(y^{u^*}\xi_T^m, X_T^b)}\right]$$

and the value of the private signals G_i with $i = 1, \ldots, N$ to a fund investor is

$$V^a \equiv CER^s - CER^u = \frac{1}{1-R}\log\left(\frac{E\left[v^a\left(y^{s^*}\xi_T^{\mathcal{G}}, X_T^b\right)\right]}{E\left[v^a\left(y^{u^*}\xi_T^m, X_T^b\right)\right]}\right). \tag{3.24}$$

Proposition 6 describes the value of private information to the fund manager and investors in the presence of asymmetric fees. Since the manager cannot completely undo the incentives implemented through non-linear contract, her excess CER is no longer independent of the contract parameters. The investor's excess CER is also affected by all the parameters.

Since the derived utility function of a fund investor (3.23) is neither concave nor convex in the normalized SPD, we cannot apply Jensen's inequality to compare the value functions U^s and U^u as in Proposition 5. The intuition that risk-aversion misalignment may also lead to a negative value of information to investors with asymmetric performance fees is confirmed in the numerical examples described in Section 3.6.

3.5 Noisy Return Forecast Timing Model

In this section, we specialize to the case of a private signal with a linear-multiplicative form. To simplify the presentation, it is assumed that r, θ^m, σ^m, π^b are constant and there is only one signal received during the investment period. Thus, $N = 1$ and $\tau_0 = 0, \tau_N = T$.

Suppose that the skilled fund manager receives a private signal which informs the future market excess return with noise. We consider a signal (3.4) with the linear multiplicative form[6]

$$G \equiv g\left(S_T^m, \zeta\right) = S_T^m \zeta \quad \text{with} \quad \zeta \equiv \exp\left(\sigma^y W_T^\zeta - \frac{1}{2}(\sigma^y)^2 T\right), \tag{3.25}$$

where W^ζ is a standard Brownian motion process, independent of the market innovation W^m. Thus, $log(\zeta) \sim N(-(\sigma^y)^2T/2, (\sigma^y)^2T)$, and $E\left[\zeta \mid \mathcal{F}_t^m\right] = 1$ for all $t \in [0, T]$. The logarithm of the signal is the cumulative local excess return of the market plus a normally distributed noise term. Smaller volatility σ^y makes the signal more informative. The inverse volatility $(\sigma^y)^{-1}$ measures the information extraction skill of the fund manager. A more-skilled manager is able to extract more precise information than a less-skilled manager. An unskilled manager does not observe the signal or has a signal with pure noise.

Corollary 1. *Suppose that θ^m and σ^m are constant and the private information filtration is $\mathcal{G}_{(\cdot)} = \mathcal{F}_{(\cdot)}^m \vee \mathcal{F}_{(\cdot)}^Y$, where $\mathcal{F}_{(\cdot)}^m$ is the public information generated by the market returns and $\mathcal{F}_{(\cdot)}^Y$ is the filtration generated by the private signal (3.25). For all $t \in [0, T)$, the conditional density of the signal is given by*

$$p_t^G(x) = \frac{\phi(d(x,t))}{x\sqrt{\Sigma_{t,T}}}$$

[6] The form is an extension of the return forecast model studied in DR section 2.1 by allowing the variance of the signal noise to increase as the timing interval T increases. Thus, the manager is not able to extract more precise information about the future market returns because of shorter investment period T.

and the PIPR is given by

$$\theta_t^{\mathcal{G}} = \mathcal{D}_t^m \log p_t^G(x)_{|x=G} = \sigma^m \left(\frac{\log G - E_t[\log G]}{VAR_t[\log G]} \right),$$

where $\phi(\cdot)$ *is the standard normal probability density distribution function and*

$$\begin{aligned} d(x,t) &= \frac{\log(x) - E_t[\log(G)]}{\sqrt{\Sigma_{t,T}}}, \\ VAR_t[\log G] &= (\sigma^m)^2(T-t) + (\sigma^y)^2 T \equiv \Sigma_{t,T}, \\ E_t[\log G] &= \left(\sigma^m \theta^m - \frac{(\sigma^m)^2 + (\sigma^y)^2}{2} \right) T + \sigma^m W_t^m. \end{aligned}$$

Corollary 1 shows that the PIPR is linear in the innovation in the log signal $\log G - E_t[\log G]$ and inversely related to the log signal's conditional variance $\Sigma_{t,T}$. The sign of the PIPR is the same as that of the innovation in the log signal. As time elapses, the informed manager learns from market realized returns and revises her assessment of risk and PIPR. For a fixed signal realization and realized gross excess return S_t^m, an increase in the skill level s raises the absolute value of the PIPR. When $\sigma^y = 0$, the informed manager has perfect foresight about future returns, the PIPR explodes as the time approaches T, and an arbitrage opportunity emerges. When the signal is uninformative $s = 0$, the variance of the noise goes to infinity and the PIPR is null.

The optimal informed investment policy with the fulcrum fees is given by (3.13). When the PIPR is positive (negative), the informed manager invests more (less) in the risky stock, and the volatility of the informed fund portfolio is greater (smaller) than that of the uninformed fund portfolio. Since the optimal informed policy and the fund return volatility are linear in PIPR, they share the same structure and properties as PIPR's . For a given innovation $\log G - E_t[\log G]$, as time passes, a manager with positive news increases the share of the risky asset in the portfolio as the variance of the forecast decreases.

Corollary 2 describes an explicit formula for the incremental value of a manager's information to a fund manager with a logarithmic utility function and fulcrum fees.

Corollary 2. *In the presence of fulcrum performance fees:*

$$F(X_T^a, X_T^b) = \alpha X_T^a + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right) \text{ with } \alpha > 0, \beta_2 \geq 0,$$

the ex ante value of the private signal G, as described in (3.25)*, to a fund manager with logarithmic utility function is*

$$V^{M,f} \equiv CER^{M,s} - CER^{M,u} = \frac{1}{2}\log\left(1 + (\sigma^m/\sigma^y)^2\right). \tag{3.26}$$

As shown by the expression (3.26), the manager's excess CER is positive and increasing in the skill level and the market volatility. A manager with greater skill level is able to extract more precise information about the future market excess returns. When the market is more volatile, the private signal is more valuable to the informed manager. The value of the private information does not depend on the public market price of risk θ^m or the timing interval T.

Corollary 3 gives an explicit formula for the value of information to a fund investor with relative risk aversion R and purely proportional fees.

Corollary 3. *Suppose managers' compensation at time T is purely proportional to the terminal value of the managed portfolio $F(X_T^a) = \alpha X_T^a$ with $\alpha > 0$ and investors' relative risk aversion coefficients $R < 1 + \sqrt{1 + (\sigma^y/\sigma^m)^2}$.* [7] *The incremental value of the log manager's private signal G (relative to public information), as described in* (3.25)*, to the investor and its first derivative with respect to the skill $s = 1/\sigma^y$ are*

$$V^p = \log\sqrt{1 + \left(\frac{\sigma^m}{\sigma^y}\right)^2} + \frac{\log\sqrt{1 - \frac{(R-1)^2(\sigma^m)^2}{(\sigma^m)^2 + (\sigma^y)^2}}}{R-1} + \frac{(R-1)^2(R-2)(\theta^m)^2 T(\sigma^m)^2}{2\left(R(R-2)(\sigma^m)^2 - (\sigma^y)^2\right)},$$

$$\frac{\partial V^p}{\partial s} = \frac{(2-R)(\sigma^m)^2\left[\left(R(\sigma^m)^2 + 1/s^2\right)\left(1 - \frac{(R-1)^2(\sigma^m)^2}{(\sigma^m)^2 + 1/s^2}\right) + (R-1)^2(\theta^m)^2 T/s^2\right]}{s\left[R(R-2)(\sigma^m)^2 - 1/s^2\right]^2}.$$

[7] This condition guarantees the ex ante expected utility of a fund investor invests in the skilled manager is finite.

When $R < 2$, the value of private information V^p is positive and it is increasing in the skill. Conversely, when $R > 2$, the value of private information V^p becomes negative and it is decreasing in the skill. Consequently, investors with relative risk aversion $R < 2$ would choose the manager with the ***highest level of skill*** *on the market, and investors with $R > 2$ would prefer the manager with the* ***lowest level of skill****.*

Corollary 3 extends the result in Proposition 5 by showing that investors could be categorized into two groups: one group who would choose the most-skilled manager and the other group who would choose the least-skilled manager. The private information is only valuable and increases in skill level for investors whose relative risk aversion is smaller than 2, the relative prudence of the log manager. These investors would choose managers with the highest skill level on the market. If the investor's relative risk aversion is greater than 2, the value of the manager's private signal to investors becomes negative due to its private nature, and the private information's negative effect on the investors' utility is more prominent as the manager's skill level increases. As a result, investors with $R > 2$ are better off choosing the least skilled manager.

We illustrate the intuition behind the clientele effect in Figure 3·2, which presents the properties of fund returns in the noisy return forecast timing model. From the perspective of an investor who has only public information, a higher skill level of manager increases the portfolio's downside tail risk as shown in the right panel. This directly follows from the fact that the manager's anticipative information is noisy. A more skilled fund manager may suffer from larger losses when her signal is misleading. This explains why investors with sufficiently high relative risk aversion will choose the least skilled fund.

The result that there exist two distinctive groups of investors is notable and has important implications. The left panel in Figure 3·2 shows that the Sharpe ratio of the skilled fund's net return is monotonically increasing in the manager's level of

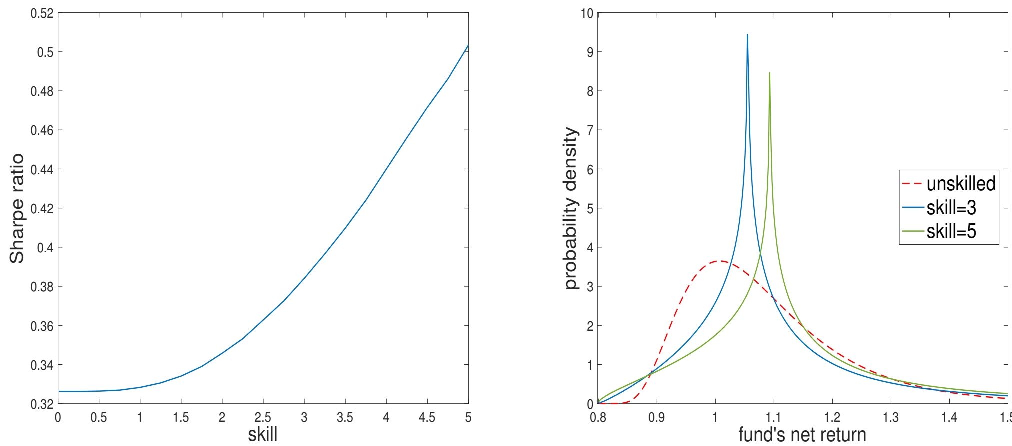

Figure 3·2: *Sharpe ratio and probability density of after-fee fund returns under public information in the noisy return forecast model.* The left panel shows the Sharpe ratios generated versus the manager's skill under the fulcrum fee contract. The right panel plots the probability density function of after-fee fund returns with three different skill levels under the fulcrum fee contract. The fixed parameter values are $\alpha = 0.6\%, \beta_1 = \beta_2 = 2\%, \sigma^m = 0.155, \theta_m = 0.47, \delta = 1, \pi^b = 0, T = 1$.

skill. Since Sharpe ratio is a commonly used criterion to consider when investors make investment decisions, Corollary 3 implies that a certain group of investors would not invest in the most skilled funds despite that a high Sharpe ratio is generated. The result thus highlights the importance of controlling heterogeneity in investors' risk preference when one evaluates investors' fund investment.

Corollary 4 gives explicit formulas for the optimal portfolio choices and fund value in the presence of asymmetric fees.

Corollary 4. *Define the constant* $\Delta_\beta = \frac{\beta_2}{1+\beta_2} - \frac{\beta_1}{1+\beta_1}$. *With asymmetric performance fees:*

$$F\left(X_T^a, X_T^b\right) = \alpha X_T^a - \alpha\beta_1 \left(X_T^a - \delta X_T^b\right)^- + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right)^+ \text{ with } \alpha > 0, \beta_2 > \beta_1 > 0,$$

the optimal fund value based on public and private information at time $t \in [0, T]$ are

$$X_t^{a,u^*} = \frac{1}{y^{u^*}\xi_t^m} + \frac{\beta_2 \delta X_t^b}{1+\beta_2}\mathcal{N}(d_{1,t}) + \frac{\beta_1 \delta X_t^b}{1+\beta_1}\mathcal{N}(-d_{1,t}), \tag{3.27}$$

$$X_t^{a,s^*} = \frac{1}{y^{s^*}\xi_t^{\mathcal{G}}} + \frac{\beta_2 \delta X_t^b}{1+\beta_2}\left(\mathcal{N}(d_{2,t}^+) - \mathcal{N}(d_{2,t}^-)\right) + \frac{\beta_1 \delta X_t^b}{1+\beta_1}(\mathcal{N}(-d_{2,t}^+) + \mathcal{N}(d_{2,t}^-)) \tag{3.28}$$

and their optimal weights invested in stock are given by

$$\pi_t^{a,u^*} = \frac{\theta^m}{\sigma^m} + \frac{\beta_2}{1+\beta_2}\frac{\delta X_t^b}{X_t^{a,u^*}}\left(\pi^b - \frac{\theta^m}{\sigma^m}\right) - \frac{\Delta_\beta \delta X_t^b}{X_t^{a,u^*}}\left(\mathcal{N}(-d_{1,t})\left(\pi^b - \frac{\theta^m}{\sigma^m}\right) + \frac{\mathcal{N}'(d_{1,t})}{\sqrt{T-t}}\right) \tag{3.29}$$

$$\begin{aligned}\pi_t^{a,s^*} &= \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m} + \frac{\beta_2}{1+\beta_2}\frac{\delta X_t^b}{X_t^{a,s^*}}\left(\pi^b - \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}\right) \\ &\quad - \frac{\Delta_\beta \delta X_t^b}{X_t^{a,s^*}}\left((\mathcal{N}(-d_{2,t}^+) + \mathcal{N}(d_{2,t}^-))\left(\pi^b - \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}\right) - (\eta_t^+ \mathcal{N}'(d_{2,t}^+) - \eta_t^- \mathcal{N}'(d_{2,t}^-))\right)\end{aligned} \tag{3.30}$$

where y^{u^*} solves $E\left[\xi_T^m X_T^{a,u^*}\right] = x^a$ and y^{s^*} solves $E\left[\xi_T^{\mathcal{G}} X_T^{a,s^*} \,\middle|\, \mathcal{G}_0\right] = x^a$, $\mathcal{N}(\cdot)$ is the standard normal cumulative distribution function and

$$d_{1,t} = \frac{\log\left(\frac{\log(\frac{1+\beta_2}{1+\beta 1})}{\xi_t^m X_t^b y^{u^*}\delta\Delta_\beta}\right) - \frac{(\pi^b\sigma^m - \theta^m)^2(T-t)}{2}}{\sqrt{(\pi^b\sigma^m - \theta^m)^2(T-t)}},$$

$$d_{2,t}^{\pm} = \frac{\pm\sqrt{\Sigma_{T,T}\left(2\log\left(\frac{\sqrt{\frac{\Sigma_{t,T}}{\Sigma_{T,T}}}\log\frac{1+\beta_2}{1+\beta 1}}{\xi_t^{\mathcal{G}} X_t^b y^{s^*}\delta\Delta_\beta}\right) + \Sigma_{t,T}\left(\pi^b - \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}\right)^2\right)} - \Sigma_{t,T}\left(\pi^b - \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}\right)}{\sqrt{\Sigma_{t,T} - \Sigma_{T,T}}}$$

$$\eta_t^{\pm} = \pm\sqrt{\frac{\frac{T(T-t)}{t^2}\left(\frac{(\sigma^m)^2 + (\sigma^y)^2}{\sigma^y}\left(\frac{\theta_t^{\mathcal{G}}}{\sigma^m} - \frac{1}{2}\right) + \frac{\sigma^m\theta^m}{\sigma^y}\right)^2}{2\log\left(\frac{\sqrt{\frac{\Sigma_{t,T}}{\Sigma_{T,T}}}\log\frac{1+\beta_2}{1+\beta 1}}{\xi_t^{\mathcal{G}} X_t^b y^{s^*}\delta\Delta_\beta}\right) + \Sigma_{t,T}\left(\pi^b - \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}\right)^2}} - \frac{1}{\sqrt{T-t}}.$$

The uninformed and informed managers' compensation at time T is

$$F(X_T^{a,u^*}, X_T^b) = \frac{\alpha(1+\beta_2)}{y^{u^*}\xi_T^m}\mathbb{1}_{\{y^{u^*}\xi_T^m \le \Psi(X_T^b)\}} + \frac{\alpha(1+\beta_1)}{y^{u^*}\xi_T^m}\mathbb{1}_{\{y^{u^*}\xi_T^m > \Psi(X_T^b)\}}, \tag{3.31}$$

$$F(X_T^{a,s^*}, X_T^b) = \frac{\alpha(1+\beta_2)}{y^{s^*}\xi_T^{\mathcal{G}}}\mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} \le \Psi(X_T^b)\}} + \frac{\alpha(1+\beta_1)}{y^{s^*}\xi_T^{\mathcal{G}}}\mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} > \Psi(X_T^b)\}}, \tag{3.32}$$

where $\Psi(b) = \log\left(\frac{1+\beta_2}{1+\beta_1}\right) / (\delta b \Delta_\beta)$.

As shown in (3.29) and (3.30), the manager's optimal portfolio is a sum of a standard mean-variance component plus additional components. The difference between the manager's optimal portfolio policy and the mean-variance demand can be interpreted as the hedging demands, motivated by the asymmetric performance fees. The second component is the hedging demand due to the performance bonus $\alpha\beta_2(X_T^a - \delta X_T^b)^+$. The last component is the hedging demand due to the performance penalty $-\alpha\beta_1(X_T^a - \delta X_T^b)^-$. Equations (3.31) and (3.32) show that in the presence of asymmetric fees the optimal trading strategies cannot fully hedge the manager's exposure to the benchmark portfolio. There is a jump in the manager's end-of-period compensation when the normalized SPD $y^*\xi_T$ hits the critical value $\Psi(X_T^b)$. As in the linear contract case, the proportional fee α does not affect the manager's optimal portfolio either. However, α and other contract parameters do affect the manager's compensation as well as the investor's after-fee wealth.

The informed agent's price of risk is changed from θ^m to $\theta^m + \theta^{\mathcal{G}}$ due to the private information. When the private signal is uninformative or the manager lacks true timing skill, $\sigma^y = \infty$, the PIPR is null and the optimal portfolio of the skilled manager as given by (3.30) collapses to that of his unskilled peer as given by (3.29).

In particular, expression (3.27) and (3.28), evaluated at $t = T$, identify the optimal fund value and can be used for the computation of the investor's and manager's CERs.

3.6 Numerical Example

This section conducts a numerical analysis of the value of information and investors' preference between skilled and unskilled funds in the noisy return forecast model. Section 3.6.1 examines the value of the manager's information to investors under the three commonly used contracts. Section 3.6.2 conducts a sensitivity analysis of the investor's relative risk aversion threshold and the key parameters.

We calibrate the parameters of risky and risk-free assets using quarterly U.S. data beginning in 1947 and ending in the first quarter of 2010. The risky asset is constructed using the CRSP value-weighted index, while the risk-free rate is constructed from real returns on the 3-month Treasury bill. The market parameters are $\theta^m = 0.47, \sigma^m = 15.5\%$, $r = 3.5\%$.

We consider three performance fee structures: purely proportional fees, fulcrum performance fees, and asymmetric performance fees. In the last two cases, the performance fee is added on top of a non-zero proportional fee ($\alpha > 0$). Based on the evidence reported by Cuoco and Kaniel (2011) that "the value-weighted average proportional component across funds charging performance fee is 60 basis, and the typical fulcrum performance fee is 2%; both on an annual basis." We set $\alpha/T = 0.6\%$ and $\beta_1 = \beta_2 = 2\%/\alpha$ for the fulcrum fees, where T is the investment horizon. For asymmetric fees, we analyze the most commonly observed two-twenty hedge fund contract and set $\alpha/T = 2\%$, $\beta_1 = 0$, and $\beta_2 = 20\%/\alpha$. For ease of exposition, we set the benchmark portfolio's weight in the stock π^b to be 0.[8] Ma et al. (2016) document "the performance evaluation window in mutual fund industry ranges from one quarter to ten years, and the average evaluation window is three years." We consider an investment horizon T of three years. The skill level $s = 1/\sigma^y$ is calibrated to be in the range of $(0, 10)$ to deliver a range of $(0, 20\%)$ for excess certainty equivalent return

[8]Hedge funds usually use 0% return or treasury rates as the benchmark in the incentive scheme (see Brown et al., 1999). The results for $\pi^b \in (0, 1]$ are qualitatively similar.

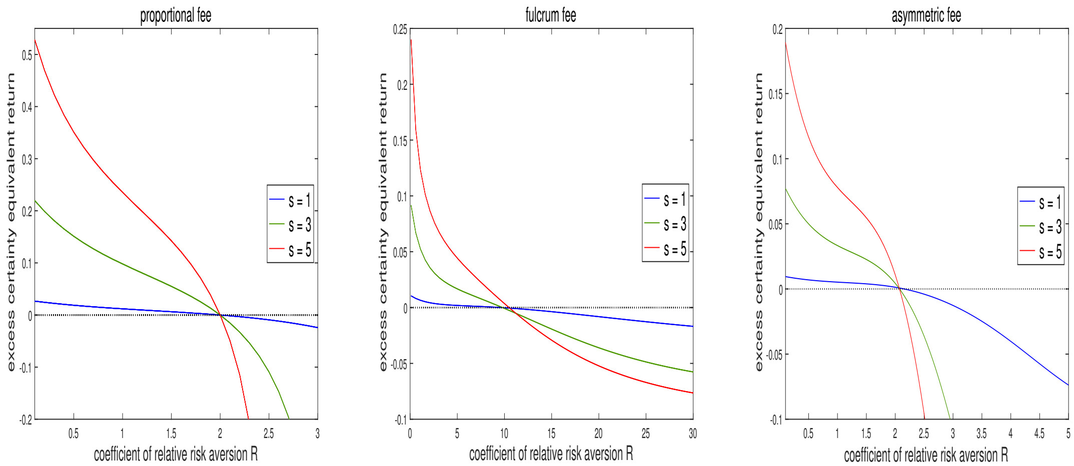

Figure 3·3: *The value of private information to investors in the noisy return forecast model.* The left panel plots the investor's excess CER from delegation under purely proportional fee contract as a function of the investor's RRA coefficient, with different skill levels $s = 1, 3, 5$. Excess CERs are computed as the investors' CER from delegation to skilled funds in excess of the CER from delegation to an unskilled fund. The contract parameters are $\alpha/T = 0.6\%, \pi^b = 0$. The middle panel plots the investor's excess CER from delegation under the fulcrum fee contract as a function of the investor's RRA coefficient, with different skill levels $s = 1, 3, 5$. The contract parameters are $\alpha/T = 0.6\%, \beta_1 = \beta_2 = 10/3, \pi^b = 0$. The right panel plots the investor's excess CER from delegation under asymmetric fee contract as a function of the investor's RRA coefficient, with different skill levels $s = 1, 3, 5$. The fixed parameters values are $r = 3.5\%, \theta^m = 0.47, \sigma^m = 0.155, \delta = 1, T = 3$.

under the setting of purely proportional fees. We set the parameter $\delta = X_0^a/X_0^b = 1$.

3.6.1 The Value of Information

Figure 3·3 presents the investor's excess CER from delegation under three different fee structures (purely proportional fees, fulcrum fees, and asymmetric fees) as a function of the investor's coefficient of relative risk aversion R, for various timing skill levels. It shows that in all three cases the investor's excess CERs are all decreasing in R and become negative as R hits a threshold value R^*.[9] It is notable that when investors

[9]The threshold value R^* is constant under purely proportional fees, while it depends on the parameters under asymmetric and symmetric fees.

are sufficiently risk-averse ($R > R^{*}$), the value of the manager's private information to investors is negative and investors prefer the unskilled fund to the skilled ones. This suggests that there are different groups of investors to skilled and unskilled funds under commonly observed portfolio management contracts. Under the same fee structure, investors in skilled funds are a more risk-tolerant clientele than investors in unskilled funds.

Notably, Figure 3·3 also shows that the higher the skill, the steeper the lines of investor's excess CERs. The value of the private information to the relatively risk-tolerant individuals ($R < R^{*}$) increases as the skill level of manager increases. On the contrary, relatively risk-averse individuals ($R > R^{*}$) suffer from larger losses when the manager is more skilled. This implies that some asset allocation or hiring decisions are inappropriate. If fund investors or owners are sufficiently more risk-averse than the managers, they should not delegate the portfolio management to the manager with a higher skill level.

The intuition for these results is as follows. Private information adds value to investors because it helps investors better assess investment opportunities. An increase in the manager's skill increases private information precision. However, the investors do not just evaluate the benefits of private information precision but also take into account the cost incurred by the noisy nature of this information. The downside tail risk of portfolios is also increasing in the manager's skill and has a negative effect on the value of private information. When the investor's risk aversion is sufficiently low, the information precision effect dominates and the investor would choose the fund manager with the highest skill. Conversely, when the investor's risk aversion is sufficiently high, the downside tail risk effect dominates and the investor would prefer the unskilled fund.

As shown in Proposition 1, and displayed in the left panel of the figure, the threshold value for the investor's coefficient of relative risk aversion is the relative prudence of the logarithmic fund manager irrespective of the manager's skill level in the purely proportional fee case. In the presence of fulcrum fees, the risk aversion threshold value is larger than that with purely proportion fees. The middle panel shows that the threshold value R^* is around 10, and investors with $R < R^*$ prefer the skilled fund and are able to extract positive value from the private signal under delegation. Conversely, investors with $R \geq 2$ will prefer the fund without any anticipative information. Interestingly, for the two-twenty asymmetric performance fees, the threshold value for the coefficient of the relative risk aversion is around 2, as displayed in the right panel. In contrast to the constant threshold in purely proportional fee case, the threshold value R^* in both the fulcrum and asymmetric fee cases is not constant and is affected by the manager's skill, the contract parameters, and the market conditions.

Table 3.1 shows investors' CER from delegation to a skilled fund manager and the value of manager's private information to the investors (excess CER) across the three commonly observed types of contracts: proportional-only fees ($\alpha > 0, \beta_1 = \beta_2 = 0$), asymmetric performance fees ($\alpha > 0, \beta_1 = 0, \beta_2 > 0$), and symmetric performance fees ($\alpha > 0, \beta_1 = \beta_2 > 0$). The excess CERs are computed as the investors' CER from delegation to the skilled funds in excess of the CER from delegation to the unskilled fund as given by (3.20) and (3.24). Table 3.1 shows that symmetric performance fee contract dominates proportional-only fee contract and asymmetric performance fee contract for investors with risk aversion larger than or equal 2 in the sense that the CER and excess CER are both higher under symmetric fee contract. For the relatively risk-averse investors with $R \geq 2$, the fact that CERs are higher indicates symmetric performance fee contract entails less welfare loss.[10] This is consistent with

[10]The welfare loss is due to the misalignment between the risk aversions of manager and investors. The manager will choose a portfolio that deviates from the investor's desired policy π^* that investors would choose if they had access to the same private information as the manager. The inclusion of symmetric performance fees might also be welfare-improving for relatively risk-tolerant investors

		CER, Excess CER (%)									
	Investor's risk aversion	1		2		3		4		5	
$s = 1$	Proportional-only	44.82	1.19	10.50	0.00	-27.34	-4.71	-76.27	-20.50	-153.34	-64.44
	Asymmetric	26.03	-1.06	-18.91	-0.17	-68.14	-3.32	-113.60	-10.59	-158.04	-20.89
	Symmetric	17.61	-0.55	15.03	0.29	13.04	0.19	11.44	0.15	10.10	0.13
$s = 3$	Proportional-only	53.42	9.79	10.50	0.00	-125.47	-102.83	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	Asymmetric	34.81	9.85	-17.21	1.54	-129.26	-64.45	-293.35	-190.34	-385.59	-251.45
	Symmetric	21.43	4.36	17.20	2.47	14.57	1.72	12.68	1.39	11.19	1.23
$s = 5$	Proportional-only	67.15	23.52	10.50	0.00	-1.45e6	-1.45e6	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	Asymmetric	48.26	23.29	-19.21	-0.46	-186.88	-122.07	-335.82	-232.81	-419.85	-282.71
	Symmetric	27.23	10.16	20.79	6.06	17.36	4.51	15.04	3.75	13.30	3.33

Table 3.1: *Investors' CER and excess CER from delegation.* This table reports the investors' certainty equivalent returns (CER) from delegation to a skilled fund manager and excess CER, which measures the value of the manager's information to the investors, under proportional-only fees, asymmetric performance fees and symmetric fees for different skill levels and investors' relative risk aversion R. Excess CER is computed as the CER from delegation to a skilled fund in excess of the CER from delegation to an unskilled fund. The fixed parameter values are $r = 0.035, \theta^m = 0.47, \sigma^m = 0.155, \delta = 1, T = 3, s = 1, \alpha/T = 0.02, \pi_b = 0$. For proportional-only fees, $\beta_1 = \beta_2 = 0$. For asymmetric fees, $\beta_1 = 0, \beta_2 = 0.2/\alpha$. For symmetric fees, $\beta_1 = \beta_2 = 0.2/\alpha$.

the finding of Sotes-Paladino and Zapatero (2017), in which they endogenize the contract parameters and show that symmetric performance fee contract is optimal and welfare-improving for investors irrespective of the investors' risk aversion relative to the manager's. Since the excess CERs measure the value of the manager's information to investors, the results suggest that highly risk-averse investors may include a symmetric performance fee in the manager's compensation to realize higher value from the manager's anticipative information.

3.6.2 Sensitivity Analysis

To analyze the sensitivity of the relative risk aversion thresholds under fulcrum fees, Figure 3·4 illustrates the effects of skill $s = 1/\sigma^y$, proportional fee α, fulcrum incentive β_2, market price of risk θ^m, and market volatility σ^m. The upper panels show that the relative risk aversion threshold is almost invariant to the proportional fee α but

with optimal contract parameters as in Sotes-Paladino and Zapatero (2017).

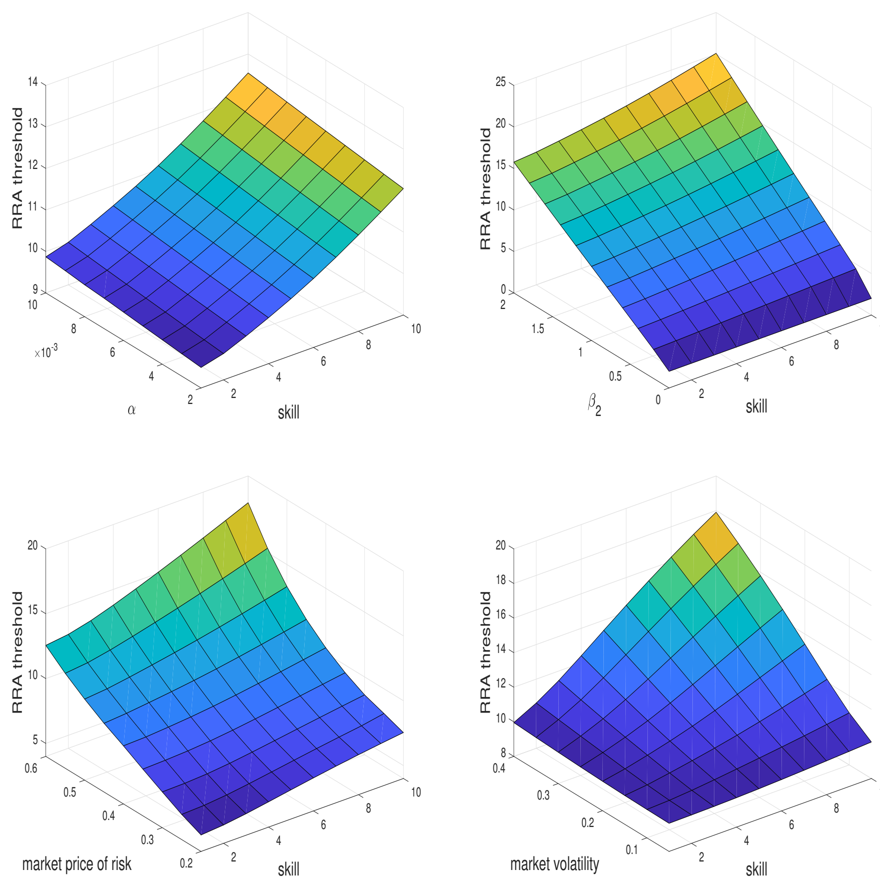

Figure 3·4: *Sensitivity of RRA threshold to model parameters under the fulcrum fee contract.* The figure presents the RRA threshold R^* of investors under fulcrum type contract versus proportional fee parameter α, performance bonus parameter β_2, (public) market price of risk θ^m, market volatility σ^m and the fund manager's skill level s. The fixed parameter values (where applicable) are $r = 3.5\%, \theta^m = 0.47, \sigma^m = 0.155, \delta = 1, T = 3, \alpha/T = 0.6\%, \beta_1 = \beta_2 = 10/9, \pi^b = 0$ and $s = 5$.

increases in the fulcrum incentive β_2. According to Proposition 5, the risk aversion threshold under fulcrum fees is larger than that under purely proportional fees. An increase in the fulcrum incentive β_2 diminishes the relative impact of proportional fee component on fund managers' compensation. Consequently, the risk aversion threshold is higher with more powerful incentive β_2. Similarly, since the relative impact of the proportional component is independent of the parameter α, there is little effect of the proportional fee α on the risk aversion threshold. The lower-left panel shows the clientele of skilled fund expands as the market improves (θ^m increases) and a one-unit increase in the skill would lead to a larger increase in the risk aversion threshold when the market price of risk is higher. This is because the negative effect of downside risk on investor's choice is alleviated in good states. By contrast, the value of the private information to a manager does not depend on the market price of risk θ^m as shown in Corollary 2. The lower-right panel of the figure displays that the set of investors in skilled fund is larger when the market is more volatile and this effect is more pronounced for high skill level. The higher market volatility the more valuable the information advantage. Therefore, more investors are investing in skilled funds. As shown in all the panels of the figure, the investor's relative risk aversion threshold is increasing in skill. The intuition is as follows. The higher the manager's level of skill, the more advantageous the private information relative to public information, the higher the value of the private information to the investors in skilled funds.

Figure 3·5 plots the risk aversion threshold of investors as a function of key parameters with asymmetric performance fees. It shows that in the absence of penalty component ($\beta_1 = 0$) the risk aversion threshold with different parameters lies in a relatively small range $[1.8, 2.3]$. This indicates that the proportional fee α, bonus incentive β_2, market price of risk θ^m, market volatility σ^m, and skill s have little impact

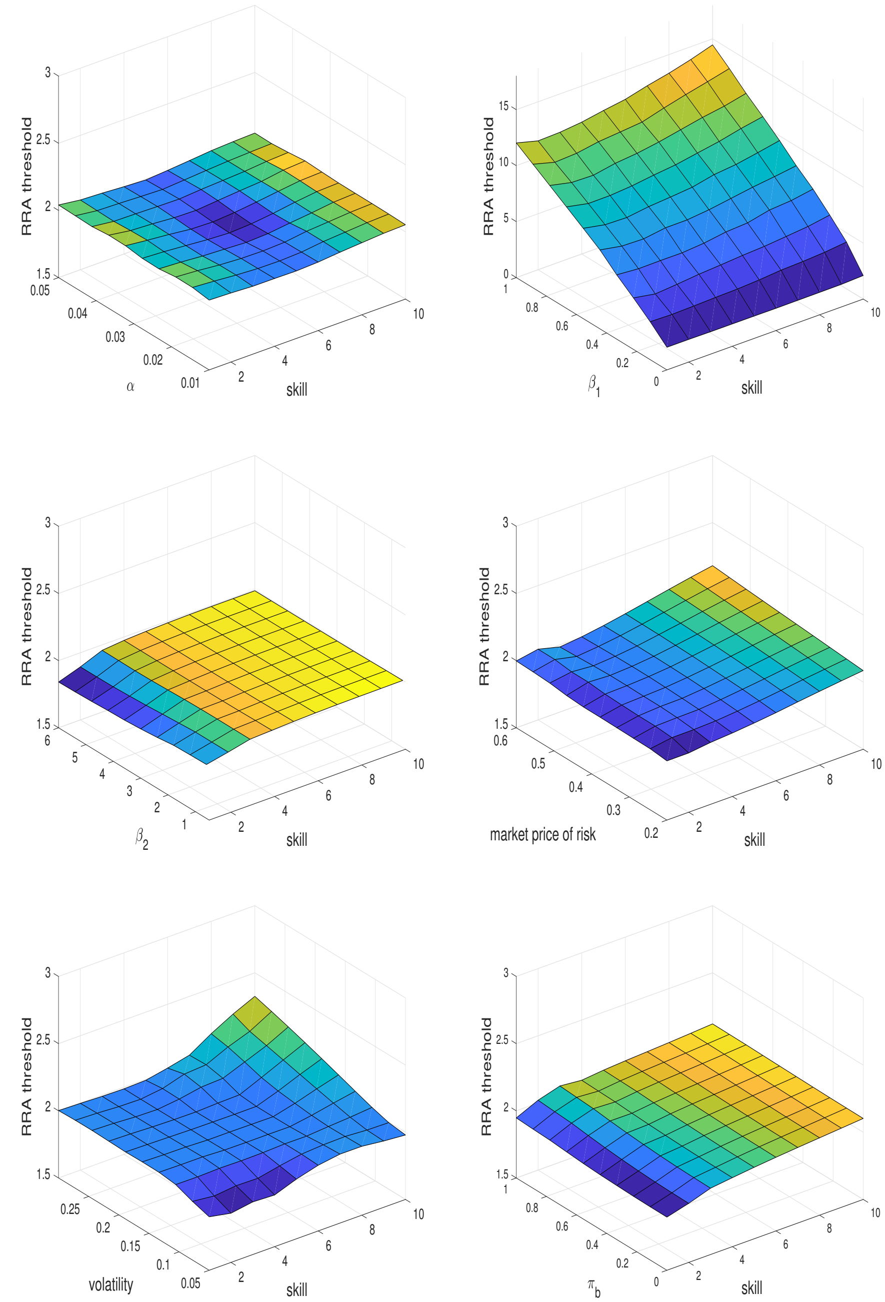

Figure 3·5: *Sensitivity of RRA threshold to model parameters under asymmetric fees contract.* The figure presents the RRA threshold R^* of investors under asymmetric performance fees contract versus proportional fee parameter α, penalty parameter β_1, performance bonus parameter β_2, (public) market price of risk θ^m, market volatility σ^m, the skill level s and the benchmark portfolio's weight in the stock π^b. The fixed parameter values (where applicable) are $r = 3.5\%, \theta^m = 0.47$, $\sigma^m = 0.155, \delta = 1, T = 3, \alpha/T = 2\%, \beta_1 = 0, \beta_2 = 10/3, \pi^b = 0$ and $s = 5$.

on the investor's preference between the skilled and unskilled fund. The risk aversion threshold seems to slightly increase in the market price of risk θ^m, market volatility σ^m, and manager's skill s. The qualitative relationships are similar to those under fulcrum fee contract. One major observation in Figure 3·5 is that the range of the risk aversion threshold as a function of the penalty sensitivity β_1 is much larger than that of other parameters. This implies that the value of market timing to investors is much more affected by the penalty sensitivity β_1 than the bonus incentive β_2 or other parameters. Increasing the penalty sensitivity β_1 leads to a wider investor clientele in skilled funds. This is because an increase in the penalty sensitivity β_1 causes the managers to reduce portfolio volatility and alleviates investors' concerns about the larger tail risk in the skilled fund relative to the unskilled fund.

The fulcrum fee contract can be regarded as the extreme case of the asymmetric fee contract with underperformance penalty sensitivity β_1 equals to the outperformance bonus incentive β_2. Comparing the Figure 3·4 and Figure 3·5 further illustrates that the qualitative impact of adding a penalty component into the manager's compensation scheme on the value of the manager's information to investors. It shows that the risk aversion threshold is larger under fulcrum fees than under option-like asymmetric fees. Furthermore, the variations in the investor's risk aversion threshold as a function of different parameters are more pronounced with fulcrum fees compared to those with option-like asymmetric fees. The results suggest the important role of the underperformance penalty component in affecting the value of the manager's private information to investors and their fund investment. When investors are sufficiently more risk-averse relative to managers, the fulcrum fee contract serves the purpose of realizing positive value from the manager's private information better than the option-like asymmetric fee contract. On the other hand, if managers are able to dictate the fee structure, the unskilled managers may abstain from including an underperformance penalty in their contracts in order to expand their clientele when soliciting funds from potential investors.

3.7 Extensions of the Model

Suppose that θ^m, σ^m, r are constant and there is one signal received by the informed manager at the inception of the investment period, namely $N = 1$, $\tau_0 = 0$, and $\tau_N = T$. For tractability, we assume that the manager observes the private signal with linear multiplicative form as described in (3.25) and takes prices as given.

3.7.1 Managers with Constant Relative Risk Aversion

We generalize the logarithmic manager assumption and consider an informed manager who has general CRRA utility with a coefficient of relative risk aversion equal to R^a. The manager maximizes the expected utility of her management fee, which is a fraction of the total asset under management at the terminal date, and solves

$$\begin{aligned} &\sup_{X_T^a \in \mathcal{G}_T} E\left[\frac{(\alpha X_T^a)^{1-R^a}}{1-R^a} \,\middle|\, \mathcal{G}_0 \right], \\ &\text{s.t.} \quad E\left[\xi_T^{\mathcal{G}} X_T^a\right] = X_0^a, X_T^a \geq 0, \end{aligned}$$

where α is the proportional fee parameter and $\xi_T^{\mathcal{G}}$ is the private state price density given by

$$\xi_T^{\mathcal{G}} = \exp\left(-\int_0^T \left(r + \frac{1}{2}\left(\theta^m + \theta_v^{\mathcal{G}}\right)^2 \right) dv - \int_0^T \left(\theta^m + \theta_v^{\mathcal{G}}\right) dW_v^{\mathcal{G}} \right).$$

The next proposition describes the manager's optimal investment strategies.

Proposition 7. *Suppose that θ^m and σ^m are constant and the private information filtration is $\mathcal{G}_{(\cdot)} = \mathcal{F}_{(\cdot)}^m \bigvee \mathcal{F}_{(\cdot)}^Y$, where $\mathcal{F}_{(\cdot)}^m$ is the public information generated by the market returns and $\mathcal{F}_{(\cdot)}^Y$ is the filtration generated by the private signal (3.25). Consider a manager with a coefficient of relative risk aversion equal to R^a. The manager's* *optimal fund value at time $t \in [0, T]$ is given by*

$$X_t^{a^*} = \left(\xi_t^{\mathcal{G}}\right)^{-1/R^a} X_0^a H_t^{\mathcal{G}} / H_0^{\mathcal{G}}$$

and the optimal risk exposure at time $t \in [0, T]$ is given by

$$\pi_t^* = \pi_t^m + \pi_t^h = \frac{\Sigma_{t,T}}{\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T}} \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}, \tag{3.33}$$

$$\pi_t^m = \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m}, \tag{3.34}$$

$$\pi_t^h = \frac{(R^a - 1)(\Sigma_{t,T} - \Sigma_{T,T})}{\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T}} \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m} \tag{3.35}$$

where

$$\begin{aligned} H_t^{\mathcal{G}} =& \sqrt{\frac{R^a \Sigma_{T,T}}{\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T}}} \left(\frac{\Sigma_{T,T}}{\Sigma_{t,T}}\right)^{-\frac{1}{2R^a}} \\ & \times \exp\left(-\left(r + \frac{\Sigma_{t,T}(\theta^m + \theta_t^{\mathcal{G}})^2}{2(\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T})}\right) \frac{(R^a - 1)(T - t)}{R^a}\right), \\ \Sigma_{t,T} =& (\sigma^m)^2 (T - t) + (\sigma^y)^2 T. \end{aligned}$$

The manager's compensation at time T is

$$F(X_T^{a^*}) = \alpha \left(\xi_T^{\mathcal{G}}\right)^{-1/R^a} X_0^a H_T^{\mathcal{G}} / H_0^{\mathcal{G}}$$

As shown by Equation (3.33), the optimal portfolio policy can be decomposed into the mean-variance demand π_t^m and the dynamic hedging demand π_t^h. By standard arguments, $\pi_t^m = \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m}$. Relative to the logarithmic case, the mean-variance demand is scaled by the manager's relative risk aversion, and the optimal portfolio choice has an additional term, reflecting the manager's hedging behavior. The dynamic hedging demand π_t^h is given by

$$\frac{d[H, W^{\mathcal{G}}]_t}{\sigma^m H_t dt} = \frac{(R^a - 1)(\Sigma_{t,T} - \Sigma_{T,T})}{\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T}} \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m} = \frac{R^a - 1}{1 + \frac{T}{T-t}\left(\frac{\sigma^y}{\sigma^m}\right)^2 R^a} \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m}. \tag{3.36}$$

Since the public market price of risk θ^m is constant, the hedging demand is motived by the stochastic fluctuations in $\theta_t^{\mathcal{G}}$. As shown in (3.36), when the manager is more risk-averse than a log manager, her hedging demand is positive (negative) if the total price of risk $(\theta^m + \theta_t^{\mathcal{G}})$ is positive (negative). Moreover, as time passes, the magnitude of the hedging demand decreases, inducing the informed manager to adjust the share of stocks in the portfolio. $\pi_t^h \to 0$ when clock approaches the terminal date T. A longer investment horizon reduces the magnitude of the hedging demand and the horizon effect is weaker for a manager with higher skill. Finally, the magnitude of the hedging demand is decreasing in the variance ratio $(\sigma^y)^2/(\sigma^m)^2$. This implies that a more skilled manager will have a larger hedging demand.

With proportional only fees, the risk sharing is perfect and the proportional fee does not affect the manager's portfolio. The proportional fee parameter α only affects the manager's compensation as well as the investor's welfare.

Corollary 5 describes the value of the private information to a fund investor (Excess CER) when both the manager and investor have CRRA under purely proportional fees.

Corollary 5. *Suppose the manager's compensation at time T is purely proportional to the terminal value of the managed portfolio $F(X_T^a) = \alpha X_T^a$ with $\alpha > 0$ and the manager (respectively investor) has CRRA utility with a coefficient of relative risk aversion equal to R^a (respectively R). The incremental value of the manager's private signal G, as described in* (3.25)*, to the investor is*

$$V^p = \begin{cases} \log\sqrt{1 + \frac{(\sigma^m)^2}{R^a(\sigma^y)^2}} + \frac{\log\sqrt{1-\frac{(R-1)(R-R^a)(\sigma^m)^2}{R^a((\sigma^m)^2+R^a(\sigma^y)^2)}}}{R-1} & \\ +\frac{(R-R^a)^2(P^a-R)(\theta^m)^2(\sigma^m)^2T}{2(R^a)^2(R(P^a-R)(\sigma^m)^2+(R^a)^2(\sigma^y)^2)}, & \text{if } R < R^e \\ -\infty, & \text{if } R \geq R^e, \end{cases}$$

where $R^e = \frac{P^a+\sqrt{(P^a)^2+4(\sigma^y/\sigma^m)^2(R^a)^2}}{2}$ and P^a is the manager's relative risk prudence coefficient.

Note that as the precision of the private signal goes to infinity, namely $\sigma^y \to 0$, the investors with relative risk aversion R smaller than the manager's relative risk *prudence P^a will choose the skilled fund. Conversely, the investors whose relative risk aversion R lies above the manager's relative risk prudence P^a will prefer the unskilled fund.*

The fact that the investors with $R > P^a$ will prefer the uninformed manager to the informed one though the informed manager has nearly perfect private information about the future market returns is remarkable. Corollary 5 implies that the clientele effect result still holds under the case of managers with CRRA preference. Assuming the manager has logarithmic utility $R^a = 1$ leads to a special case of the finding in Corollary 3.

Figure 3·6 illustrates the threshold of investor's RRA when the managers also have general CRRA utility. The threshold values can be obtained by finding the root of $V^p = 0$ numerically. R^e provides an upper bound for the RRA threshold. Contrary to the logarithmic manager case, the investor's RRA thresholds here are affected by the manager's skill $s = 1/\sigma^y$ as well as the market conditions θ^m, σ^m. The upper left panel shows that the investor's RRA threshold is decreasing in the manager's skill when the manager's RRP is above 2, whereas the investor's RRA threshold is slightly increasing in the manager's skill level when the manager's RRP lies below than 2. Consistent with the theoretical finding in Corollary 5, it appears that the investor's RRA threshold converges to the manager's relative risk prudence as the skill level increases. The remaining panels show that the investor's RRA threshold is decreasing in both the market price of risk and market volatility when the manager's RRP is above 2. However, the pattern changes when the manager's RRP is below 2.

3.7.2 Passive Alternative

We further extend the model in Section 3.7.1 by allowing the fund investors to choose a passively managed index fund (e.g., market index) as an alternative to the actively managed funds. Suppose that investors choose among an actively managed skilled

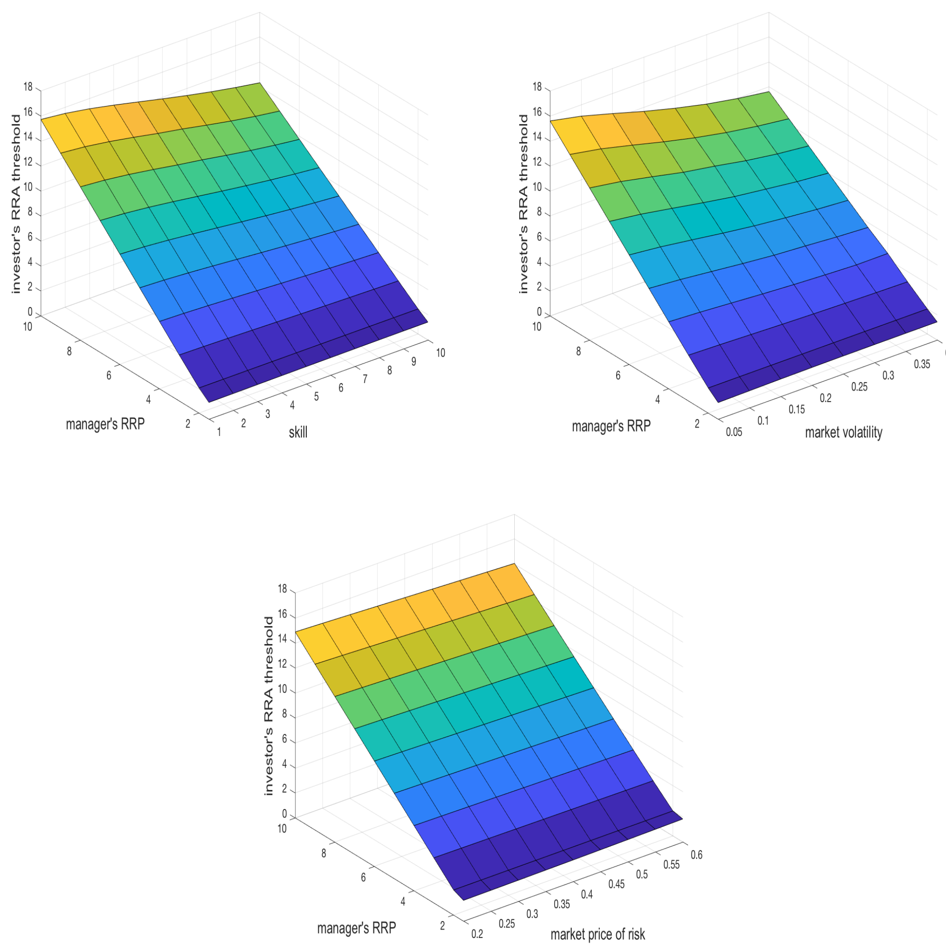

Figure 3·6: *Investor's RRA threshold versus manager's RRP under purely proportional fees.* The figure presents the relative risk aversion threshold of investors under purely proportional fee contract versus manager's relative risk prudence coefficients and skill levels. The fixed parameter values (where applicable) are $r = 3.5\%, \theta^m = 0.47, \sigma^m = 0.155, s = 5, T = 3, \alpha/T = 2\%$.

fund, an actively managed unskilled fund, and the market index. The fund investor's portfolio-choice problem can be characterized as

$$\max_{\mathscr{C}\in\{s,u,m\}} E\left[\frac{\left((1-\alpha^{\mathscr{C}})X_T^{\mathscr{C}}\right)^{1-R}}{1-R}\right], \tag{3.37}$$

where X_T^s (respectively X_T^u) is the optimal terminal fund value chosen by the skilled (respectively unskilled) fund manager with relative risk aversion R^a and X_T^m is the market index as given by Equation (3.1). The parameter $\alpha^{\mathscr{C}}$ with $\mathscr{C} \in \{s,u,m\}$ represents the proportional fee charged by these funds.

Figure (3·7) illustrates the fund investor's choice among a skilled fund manager, an unskilled fund manager, and the market index as a function of the investor's RRA and the manager's RRP. It highlights the impact of the manager's and investor's risk preference on fund investors' investment behaviors.

Although some investors turn to index investing, the clientele effect still exists in the active management industry. Unskilled fund investors (represented by the dark blue area) are generally more risk-averse than skilled fund investors (represented by the light blue area). Figure (3·7) illustrates this finding for different level of skills. A bit surprisingly, the unskilled fund is not completely dominated by the passive index fund for relatively risk averse investors, but depending on the RRP of the fund managers. When the fund managers are relatively risk-averse (e.g., $P^a > 5$), the investors whose risk aversion above the red solid line will choose the unskilled fund manager. The intuition is that as the manager's risk aversion increases, the misalignment between the risk preference of the unskilled manager and the relatively risk-averse investors becomes smaller than that between the risk exposure of the market index and the investors' optimal portfolios. In the same vein, when the risk preference misalignment between manager and investors $|R - R^a|$ is large enough, the alignment of the passive fund and the investors' optimal portfolios is better than

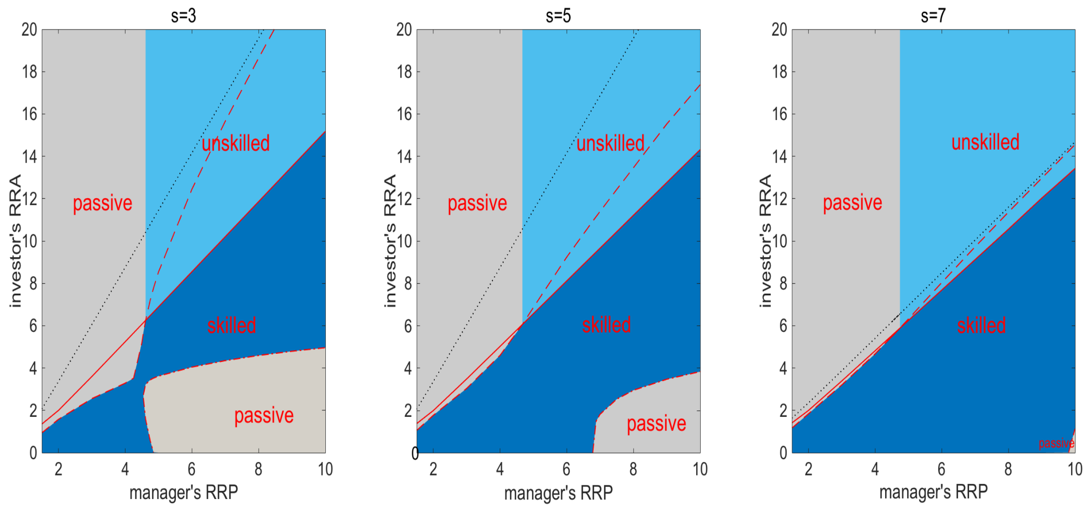

Figure 3·7: *Investor's choice among skilled, unskilled, and index funds under purely proportional fees.* The figure presents the investor's choice among skilled, unskilled, and index funds under purely proportional fee contract. The dark blue area represents the region in which the investors will choose skilled fund. The light blue area represents the region in which the investors will choose unskilled fund. The grey areas represent the regions in which the investors will choose the index fund. The black dotted line represents R^e as a function of manager's RRP. Parameter values are $r = 3.5\%, \theta^m = 0.47, \sigma^m = 0.155, T = 3, \alpha/T = 2\%$ for active (skilled and unskilled) funds and $\alpha/T = 1\%$ for the passively managed index fund.

that of the active fund managers' and investors' optimal portfolios. As a result, these investors will choose the passively managed index fund, as represented by the grey regions in the above plots.

Another insight derived from Figure 3·7 is that investor's choice between active and passive investing does not only depend on risk preference of manager and investor, but also on the manager's skill level. As seen in the figure, the grey regions expand as the manager's skill s decreases from 7 to 3. It implies that the investors in passively managed index funds are growing as the active managers become less skilled. This result is consistent with the findings of many studies that managers' abilities to beat the market declines as the active management industry size increases in recent decades and this could help explain the growing popularity of index funds (e.g., Berk and

Green, 2004; Pástor and Stambaugh, 2012).

3.8 Conclusion

The clientele effects in the money management industry have been widely documented. Prior literature attributes these clientele effects to irrationality or psychological tendencies. In this chapter, we establish a rational theory to explain the clientele effect in the money management industry and show that investors in skilled funds are uniformly more risk-tolerant than investors in unskilled funds.

Taking a general parametric class of contracts as given, we first derive and analyze the optimal trading strategies of the skilled fund manager who is endowed with private information about future market returns. Then we analyze the value of the private information to both managers and investors. Though the privation information is always valuable to the manager, it might not add value and may be harmful to the investors who are sufficiently more risk-averse than the manager. Investors with risk aversion exceeding a threshold value will never find it beneficial to delegate the management of their wealth to the skilled fund manager. As a result, the fund investor clientele is endogenously segmented. The relatively risk-tolerant investors will prefer skilled funds, whereas the highly risk-averse investors will prefer unskilled funds. This result provides theoretical justification for some recent empirical findings of the clientele effect in the mutual fund industry.

In the absence of performance fees, the relative risk aversion threshold that separates investors in skilled and unskilled funds equals the log manager's RRP, irrespective of the skill or other parameters. When the fund manager receives performance-based fees, the relative risk aversion threshold is affected by the skill level, contract parameters, and market conditions. We specialize the general results to a parametric timing model. A comparative static analysis of the risk aversion threshold is carried out to analyze the impacts of symmetric and asymmetric performance fees on investors' fund investment. The results suggest that sufficiently risk-averse investors should include a fulcrum performance fee in the manager's compensation contract for the purpose of realizing positive value from the manager's private information. The intuition behind this is that lifting the penalty sensitivity to the same level of bonus incentive reduces the portfolio's volatility and alleviates highly risk-averse investors' concerns about the larger downside tail risk in the skilled fund relative to the unskilled fund.

Our qualitative results do not depend on the assumptions that managers have logarithmic utility function or investors are restricted to choose among actively managed funds. Extensions to the basic setup examine the cases of managers with general CRRA utility and investors who can invest in a passive alternative. There still exist two distinctive groups of investors for skilled and unskilled funds. Moreover, our results in Section 3.2 are easily generalizable to a multi-asset setting that managers have both the timing and selection skills.

Chapter 4

Conclusions

This dissertation examines the dynamic asset allocation problem in the presence of regime switching and provides a rationale for the clientele effect in the money management industry. The first model accounts for the regime switching dynamics in the asset returns and a dynamic asset allocation problem is solved under a Bayesian framework. The second model provides a theoretical justification for the clientele effect in the money management industry and we study the implications of portfolio management contracts on fund investment decisions.

In the first model. the results show that regimes not only exist in systematic components of sector ETFs characterized by a common state variable in risk factors and factor loading but also in idiosyncratic return volatilities. The idiosyncratic volatilities exhibit similar regime cycles across different sectors. A new regime switching multi-factor model is proposed to capture the different dynamics in the systematic and idiosyncratic components of asset returns. A Gibbs sampling method is introduced to deal with the computational challenge due to the introduction of multiple Markov Chains and a large number of risky assets.

The empirical asset allocation experiments confirm the significant economic gains of accounting for regimes in portfolio decisions. Investors change their portfolio weights considerably over time as they recursively update their beliefs about the underlying state probabilities and tend to hold more cash in high volatile states. Moreover, the out-of-sample experiments show that the proposed regime switching model outperforms the traditional regime switching multi-factor models in terms of certainty equivalent return, especially at the short horizons. This highlights that correct specification of regime structure and number of regimes are of equal importance in asset allocation as accounting for regimes.

In the second model, the fund manager's skill comes from his anticipative information from future market returns and the private nature of this information can be costly and even adverse to fund investors. Investors with risk aversion exceeding a threshold value will never find it beneficial to delegate the management of their wealth to the skilled fund manager. The fund investor clientele is thus endogenously segmented. Investors in the skilled fund are uniformly more risk-tolerant than investors in the unskilled funds.

We also examine the impacts of commonly observed portfolio management contracts on the portfolio decisions of fund managers and investors. With only proportional management fees, the relative risk aversion threshold that separates investors in skilled and unskilled funds equals to the log manager's RRP, irrespective of the skill or other parameters. In the presence of performance-based fees, the relative risk aversion threshold is affected by the skill level, contract parameters, and market conditions. The comparative analysis suggests that sufficiently risk-averse investors should include a fulcrum performance fee in the manager's compensation contract for the purpose of realizing positive value from the manager's private information. The intuition behind this is that lifting the penalty sensitivity to the same level of bonus incentive reduces the portfolio's volatility and alleviates highly risk-averse investors' concerns about the larger downside tail risk in the skilled fund relative to the unskilled fund.

Appendix A

Proofs

A.1 Chapeter 3: A Rationale for the Clientele Effect in Money Management

We assume that the following regularity conditions hold as in DR.

Assumption

Let $\theta_v \equiv \theta_v^m + \theta_v^{\mathcal{G}}$ be an informed agent's price of risk. For $i = 1, \ldots, N$, the private information price of risk $\theta^{\mathcal{G}}$ and the informed trading strategy π satisfy the conditions

(i) (Finite quadratic variation) $\int_{\tau_{i-1}}^{\tau_i} \theta_v^2 dv < \infty$ P-a.s.

(ii) (Gains from trade bounded below) Let $H_i^\epsilon \equiv \frac{1}{\epsilon} \int_{\tau_{i-1}}^{\tau_i} \pi_v (R_{v,v+\epsilon}^m - E[R_{v,v+\epsilon}^m | \mathcal{G}_v]) dv$ be the innovations in the gains from trade. Then, $H_i^\epsilon > -\underline{H}_i$ for some positive $\mathcal{G}_0$-measureable random variable H_i with $E[\underline{H}_i | \mathcal{G}_0] < \infty$ P-a.s.

The first condition ensures that the private information price of risk is finite, ruling out arbitrage opportunities. The second condition rules out the "doubling strategies".

Proof of Lemma 1 . Suppose $X_b > 0$ and $\alpha > 0$, $\beta_2 > \beta_1 \geq 0$, there exist unique numbers $X_1(X_b)$ and $X_2(X_b)$ solve the system of equations

$$\begin{cases} \frac{u^M(F(X_2(X_b),X_b))-u^M(F(X_1(X_b),X_b))}{X_1(X_b)-X_2(X_b)} = u_x^M(F(X_2(X_b),X_b)\alpha(1+\beta_2), \\ u_x^M(F(X_1(X_b),X_b))(1+\beta_1) = u_x^M(F(X_2(X_b),X_b))(1+\beta_2). \end{cases} \tag{A.1}$$

In particular, if marginal utility is homogeneous of degree $-R$ $(R \neq 1)$, letting $\eta \equiv \left(\frac{1+\beta_2}{1+\beta_1}\right)^{1-1/R}$ and $\rho \equiv \left(\frac{1+\beta_2}{1+\beta_1}\right)^{-1/R}$. In particular, if marginal utility is homogeneous of degree $-R$, the above equations imply that

$$\begin{cases} X_1(X_b) + \beta_1(X_1(X_b) - \delta X_b) = \rho(X_2(X_b) + \beta_2(X_2(X_b) - \delta X_b)), \\ X_2(X_b) = X_1(X_b) + \frac{u^M(\alpha X_2(X_b)+\alpha\beta_2(X_2(X_b)-\delta X_b)-u^M(\rho(\alpha X_2(X_b)+\alpha\beta_2(X_2(X_b)-\delta X_b)))}{u_x^M(\alpha X_2(X_b)+\alpha\beta_2(X_2(X_b)-\delta X_b))\alpha(1+\beta_2)}. \end{cases}$$

Equivalently, u^M is homogeneous of degree $1-R$, we have

$$\begin{aligned} X_2(X_b) &= X_1(X_b) + (1-\rho^{1-R})\frac{u^M(\alpha X_2(X_b) + \alpha\beta_2(X_2(X_b) - \delta X_b))}{u_x^M(\alpha X_2(X_b) + \alpha\beta_2(X_2(X_b) - \delta X_b))\alpha(1+\beta_2)} \\ &= X_1(X_b) + \frac{1-\eta}{(1-R)\alpha(1+\beta_2)}(\alpha X_2(X_b) + \alpha\beta_2(X_2(X_b) - \delta X_b)), \end{aligned}$$

Direct computation yields

$$\begin{cases} X_1(X_b) = \left(\frac{(\frac{\eta}{R}-1)\frac{\beta_1}{1+\beta_1}+\eta(1-\frac{1}{R})\frac{\beta_2}{1+\beta_2}}{\eta-1}\right)\delta X_b, \\ X_2(X_b) = X_1(X_b) + \frac{1}{R}\left(\frac{\beta_2}{1+\beta_2} - \frac{\beta_1}{1+\beta_1}\right)\delta X_b. \end{cases}$$

Moreover, $\underline{X}(X_b) = \frac{\beta_1 \delta X_b}{1+\beta_1} < X_1(X_b) < \delta X_b < X_2(X_b)$. In particular, for logarithmic utility

$$\begin{aligned} X_1(X^b) &= \left(log\left(\frac{1+\beta_2}{1+\beta_1}\right)\right)^{-1}\left(\frac{\beta_2}{1+\beta_2} - \frac{\beta_1}{1+\beta_1}\right)\delta X^b + \frac{\beta_1}{1+\beta_1}\delta X^b > \underline{X}(X^b), \\ X_2(X^b) &= X_1(X^b) + \left(\frac{\beta_2}{1+\beta_2} - \frac{\beta_1}{1+\beta_1}\right)\delta X^b. \end{aligned}$$

□

Proof of Lemma 2. This closely follows the proof of Lemma 2 from Cuoco and Kaniel (2011). The first equation in the system (A.1) shows that $v^M(\cdot, X_b)$ is continuous at

$X_2(X_b)$, while the second equation in the system (A.1) show that $v^M(\cdot, X_b)$ is continuously differentiable at $X_1(X_b)$. Thus $v^M(\cdot, X_b)$ is continuously differentiable and concave on $[\underline{X}(X_b), \infty)$. Since $v^M(X_1(X_b), X_b) = u^M(F(X_1(X_b), X_b))$ and $u^M(F(\cdot, X_b))$ is strictly concave on the interval $(X_1(X_b), \delta X_b]$ while $v^M(\cdot, X_b)$ is linear, we must have that $v^M(\cdot, X_b) > u^M(F(\cdot, X_b))$ on the interval. Similarly, we have $v^M(\cdot, X_b) > u^M(F(\cdot, X_b))$ on the interval $[\delta X_b, X_2(X_b))$. Moreover, $v^M(\cdot, X_b) = u^M(F(\cdot, X_b))$ on $A(X_b)$, thus we have $v^M(\cdot, X_b) \geq u^M(F(\cdot, X_b))$ on the interval $[\underline{X}(X_b), \infty)$.

Suppose $\hat{v}^M(\cdot, X_b)$ is any concave function with $\hat{v}^M(\cdot, X^b) \geq u^M(F(\cdot, X^b))$ on the interval $[\underline{X}(X^b), \infty)$. It follows from the definition that $\hat{v}^M(\cdot, X^b) \geq u^M(F(\cdot, X^b)) = v^M(\cdot, X^b)$ on $A(X^b)$. In addition, $\hat{v}^M(\cdot, X^b)$ is concave on the interval $(X_1(B), X_2(B))$ while $v^M(\cdot, X^b)$ is linear, it follows that $\hat{v}^M(\cdot, X^b) > v^M(\cdot, X_b)$ on $(X_1(X_b), X_2(X_b))$. Thus, $v^M(\cdot, X^b)$ is the smallest concave function with $\hat{v}^M(\cdot, X^b) \geq u^M(F(\cdot, X_b))$ on the interval $[\underline{X}(X_b), \infty)$. □

Proof of Proposition 1. Suppose that $\alpha > 0, \beta_1 = \beta_2 \geq 0$. The first order condition of the fund manager's static problem is

$$\alpha(1+\beta_2)u_x^M(\alpha X_T^{a,s^*} + \alpha\beta_2(X_T^{a,s^*} - \delta X_T^b)) = y^{s^*}\xi_T^{\mathcal{G}}.$$

Direct computation yields

$$X_T^{a,s^*} = \frac{1}{\alpha(1+\beta_2)} I^M\left(\frac{y^{s^*}\xi_T^{\mathcal{G}}}{\alpha(1+\beta_2)}\right) + \frac{\beta_2\delta X_T^b}{1+\beta_2}, \tag{A.2}$$

where $I^M(\cdot)$ is the inverse function of $u_x^M(\cdot)$ and y^{s^*} is determined by the static budget constraint

$$\frac{x^a}{1+\beta_2} = E\left[\frac{\xi_T^{\mathcal{G}}}{\alpha(1+\beta_2)} I^M\left(\frac{y^{s^*}\xi_T^{\mathcal{G}}}{\alpha(1+\beta_2)}\right)\middle|\mathcal{G}_0\right] \equiv \chi(y^{s^*}).$$

The function $\chi(y)$ is continuous and strictly decreasing on $(0, \infty)$. Moreover, $\chi(y) \to 0$ as $y \to \infty$ and $\chi(y) \to \infty$ as $y \to 0$. Therefore, there exists a unique $y^{s^*} > 0$ such that $\chi(y^{s^*}) = x^a/(1+\beta_2)$. In particular, with $u^M(x) = \log(x)$, $y^{s^*} = (1+\beta_2)/x^a$. We can find the manager's optimal portfolio choice by applying Itô's Lemma on both sides of (A.2) and matching the coefficients in front of $dW_v^{\mathcal{G}}$. The optimal portfolio choice is given by

$$\pi_v^{a,s^*} = \frac{\theta_v^m + \theta_v^{\mathcal{G}}}{\sigma_v^m} + \frac{\beta_2}{1+\beta_2}\frac{\delta X_v^b}{X_v^{a,s^*}}\left(\pi^b - \frac{\theta_v^m + \theta_v^{\mathcal{G}}}{\sigma_v^m}\right).$$

Substituting the optimal fund's end-of-period value (A.2) into the contract yields the manager's compensation at time T, $F(X_T^{a,s^*}, X_T^b) = \alpha x^a / \xi_T^{\mathcal{G}}$. □

Proof of Proposition 2. Suppose that $\alpha > 0$, $\beta_2 > \beta_1 \geq 0$. The concavified utility function v^M concave and continuously differentiable on the interval $[\underline{X}(X_b), \infty)$. Thus we can solve the concavified problem using standard optimization theory. The sufficient and necessary condition for X_T^{a,s^*} to be optimal is

$$\frac{\partial v^M}{\partial x^a}(X_T^{a,s^*}, X_T^b) = y^{s^*} \xi_T^{\mathcal{G}}.$$

We can define a function $g^I : (0,\infty) \times (0,\infty)$

$$g^I(y,b) = \begin{cases} \frac{1}{\alpha(1+\beta_2)} I^M\left(\frac{y}{\alpha(1+\beta_2)}\right) + \frac{\beta_2 \delta b}{1+\beta_2} > X_2(b) & \text{if} \quad y \leq \Psi(b), \\ \frac{1}{\alpha(1+\beta_1)} I^M\left(\frac{y}{\alpha(1+\beta_1)}\right) + \frac{\beta_1 \delta b}{1+\beta_1} < X_1(b) & \text{if} \quad y > \Psi(b), \end{cases}$$

where $I^M(\cdot)$ is the inverse function of $u_x^M(\cdot)$ and $\Psi(b) = \frac{\log\left(\frac{1+\beta_2}{1+\beta_1}\right)}{\delta b\left(\frac{\beta_2}{1+\beta_2} - \frac{\beta_1}{1+\beta_1}\right)}$. The optimal end-of-period fund value is given by

$$\begin{aligned} X_T^{a,s^*} &= g^I\left(y^{s^*}\xi_T^{\mathcal{G}}, X_T^b\right) \\ &= \frac{1}{\alpha(1+\beta_2)} I^M\left(\frac{y^{s^*}\xi_T^{\mathcal{G}}}{\alpha(1+\beta_2)}\right) + \frac{\beta_2 \delta X_T^b}{1+\beta_2} \mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} \leq \Psi(X_T^b)\}} + \frac{\beta_1 \delta X_T^b}{1+\beta_1} \mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} > \Psi(X_T^b)\}}, \end{aligned}$$

where y^{s^*} is the Lagrange multiplier solving the static budget constraint:

$$x^a = E\left[\xi_T^{\mathcal{G}} g^I\left(y^{s^*}\xi_T^{\mathcal{G}}, X_T^b\right) \Big| \mathcal{G}_0\right] \equiv \tilde{\chi}(y^{s^*}).$$

The function $\tilde{\chi}(y)$ is continuous and strictly decreasing on $(0,\infty)$. Moreover, $\tilde{\chi}(y) \to \beta_1 X_0^a/(1+\beta_1)$ as $y \to \infty$ and $\chi(y) \to \infty$ as $y \to 0$. Therefore, there exists a unique $y^{s^*} > 0$ such that $\chi(y^{s^*}) = X_0^a$. With $u^M(x) = \log(x)$, we immediately obtain (3.16) and the manager's compensation at time T:

$$F(X_T^{a,s^*}, X_T^b) = \frac{\alpha(1+\beta_2)}{y^{s^*}\xi_T^{\mathcal{G}}} \mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} \leq \Psi(X_T^b)\}} + \frac{\alpha(1+\beta_1)}{y^{s^*}\xi_T^{\mathcal{G}}} \mathbb{1}_{\{y^{s^*}\xi_T^{\mathcal{G}} > \Psi(X_T^b)\}}.$$

□

Proof of Proposition 3. Define $H_v^\epsilon \equiv \frac{1}{\epsilon}\int_v^{v+\epsilon} dW_v^m$. For $v \in [\tau_{i-1}, \tau_i)$, using Bayes' rule $\frac{P(dW^m|\mathcal{F}_v,G_i=z)}{P(dW^m|\mathcal{F}_v)} = \frac{P(G_i\in dz|\mathcal{F}_{v+\epsilon})}{P(G_i\in dz|\mathcal{F}_v)}$ gives

$$\begin{aligned}
E\left[H_v^\epsilon \,|\, \mathcal{G}_v\right] =& E\left[\frac{1}{\epsilon}\int_v^{v+\epsilon} dW_s^m \,\middle|\, \mathcal{F}_v, G_i = z\right]_{|z=G_i} \\
=& E\left[H_v^\epsilon \frac{P(dW^m|\mathcal{F}_v, G_i = z)}{P(dW^m|\mathcal{F}_v)} \,\middle|\, \mathcal{F}_v\right]_{|z=G_i} \\
=& E\left[H_v^\epsilon \frac{P(G_i \in dz|\mathcal{F}_{v+\epsilon})}{P(G_i \in dz|\mathcal{F}_v)} \,\middle|\, \mathcal{F}_v\right]_{|z=G_i}.
\end{aligned}$$

Therefore, with $p_v^G(z) = P(G_i \in dz|\mathcal{F}_v)$, for $v \in [\tau_{i-1}, \tau_i)$ when $z = G_i$

$$\begin{aligned}
\theta_v^{\mathcal{G}} &= \lim_{\epsilon\downarrow 0} E\left[H_v^\epsilon \,|\, \mathcal{G}_v\right] = \lim_{\epsilon\downarrow 0} E_v\left[\frac{p_{v+\epsilon}^G(z)}{p_v^G(z)}\int_v^{v+\epsilon} dW_s^m\right]_{|z=G_i} \\
&= \left(\frac{d[\log p_v^G(z), W^m]_v}{dv}\right)_{|z=G_i} = \mathcal{D}_v \log p_v^G(z)_{|z=G_i},
\end{aligned}$$

where $E_v[\cdot] \equiv E[\cdot|\mathcal{F}_v]$.

In the presence of fulcrum performance fees $\alpha X_T^a + \alpha\beta_2\left(X_T^a - \delta X_T^b\right)$ with $\alpha > 0$ and $\beta_2 \geq 0$, the managers compensation is $\alpha x^a/\xi_T$. Thus, the ex ante value functions for the skilled and unskilled managers are

$$\begin{aligned}
E\left[\log\frac{\alpha x^a}{\xi_T^{\mathcal{G}}}\right] &= \log\left(\alpha X_0^a\right) + E\left[\int_0^T r_v dv\right] + \frac{1}{2}E\left[\int_0^T \left(\theta_v^m + \theta_v^{\mathcal{G}}\right)^2 dv\right] \\
E\left[\log\frac{\alpha x^a}{\xi_T^m}\right] &= \log\left(\alpha X_0^a\right) + E\left[\int_0^T r_v dv\right] + \frac{1}{2}E\left[\int_0^T \left(\theta_v^m\right)^2 dv\right].
\end{aligned}$$

Using $E[\theta_v^m\theta_v^{\mathcal{G}}] = 0$, which follows from $E_v[\theta_v^{\mathcal{G}}] = E_v\left[\lim_{\epsilon\downarrow 0} E\left[H_v^\epsilon \,|\, \mathcal{G}_v\right]\right] = 0$, gives $E\left[\log\frac{\alpha X_0^a}{\xi_T^{\mathcal{G}}}\right] = E\left[\log\frac{\alpha X_0^a}{\xi_T^m}\right] + \frac{1}{2}E\left[\int_0^T\left(\theta_v^{\mathcal{G}}\right)^2 dv\right]$ and the ex ante value of information to the manager

$$V^{M,f} \equiv CER^{M,s} - CER^{M,u} = \frac{1}{2}\int_0^T E\left[\left(\theta_v^{\mathcal{G}}\right)^2\right] dv.$$

The process $p_v^G(z)$ is a martingale. By the Clark-Ocone formula,

$$\begin{aligned}
p_{\tau_i}^G(z) &= p_{\tau_{i-1}}^G(z) + \int_{\tau_{i-1}}^{\tau_i} E_v[\mathcal{D}_v p_{\tau_i}^G(z)]dW_v^m = p_{\tau_{i-1}}^G(z) + \int_{\tau_{i-1}}^{\tau_i} \mathcal{D}_v E_v[p_{\tau_i}^G(z)]dW_v^m \\
&= p_{\tau_{i-1}}^G(z) + \int_{\tau_{i-1}}^{\tau_i} \mathcal{D}_v p_v^G(z)dW_v^m = p_{\tau_{i-1}}^G(z) + \int_{\tau_{i-1}}^{\tau_i} p_v^G(z)\mathcal{D}_v \log p_v^G(z)dW_v^m \\
&= p_{\tau_{i-1}}^G(z) + \int_{\tau_{i-1}}^{\tau_i} p_v^G(z)\theta_v^{\mathcal{G}}(z)dW_v^m.
\end{aligned}$$

Solving this linear stochastic differential equation gives

$$\frac{p_{\tau_i}^G(z)}{p_{\tau_{i-1}}^G(z)} = \exp\left(\int_{\tau_{i-1}}^{\tau_i} \theta_v^{\mathcal{G}}(z)dW_v^m - \int_{\tau_{i-1}}^{\tau_i} \frac{1}{2}\theta_v^{\mathcal{G}}(z)^2 dz\right). \tag{A.3}$$

and

$$\begin{aligned}
&E_{\tau_{i-1}}\left[\log \frac{p_{\tau_i}^G(z)}{p_{\tau_{i-1}}^G(z)} \middle| G_i = z\right] \\
=&E_{\tau_{i-1}}\left[\int_{\tau_{i-1}}^{\tau_i} \theta_v^{\mathcal{G}}(z)dW_v^m - \int_{\tau_{i-1}}^{\tau_i} \frac{1}{2}\theta_v^{\mathcal{G}}(z)^2 dz \middle| G_i = z\right] \\
=&E_{\tau_{i-1}}\left[\int_{\tau_{i-1}}^{\tau_i} \theta_v^{\mathcal{G}}(z)(dW_v^m - \theta_v^{\mathcal{G}}(z)dv) + \int_{\tau_{i-1}}^{\tau_i} \frac{1}{2}\theta_v^{\mathcal{G}}(z)^2 dv \middle| G_i = z\right] \\
=&\frac{1}{2}E_{\tau_{i-1}}\left[\int_{\tau_{i-1}}^{\tau_i} \theta_v^{\mathcal{G}}(z)^2 dv \middle| G_i = z\right]
\end{aligned}$$

Using Bayes' rule $\frac{P(dW^m|\mathcal{F}_{\tau_{i-1}}, G_i = z)}{P(dW^m|\mathcal{F}_{\tau_{i-1}})} = \frac{P(G_i \in dz|\mathcal{F}_{\tau_i})}{P(G_i \in dz|\mathcal{F}_{\tau_{i-1}})} = \frac{p_{\tau_i}^G(z)}{p_{\tau_{i-1}}^G(z)}$ gives

$$\begin{aligned}
\frac{1}{2}E_{\tau_{i-1}}\left[\int_{\tau_{i-1}}^{\tau_i} \left(\theta_v^{\mathcal{G}}\right)^2 dv\right] &= \int_{-\infty}^{\infty} \frac{1}{2}E_{\tau_{i-1}}\left[\int_{\tau_{i-1}}^{\tau_i} \theta_v^{\mathcal{G}}(z)^2 dv \middle| G_i = z\right] p_{\tau_{i-1}}^G(z)dz \\
&= \int_{-\infty}^{\infty} E_{\tau_{i-1}}\left[\log \frac{p_{\tau_i}^G(z)}{p_{\tau_{i-1}}^G(z)} \middle| G_i = z\right] p_{\tau_{i-1}}^G(z)dz \\
&= \int_{-\infty}^{\infty} E_{\tau_{i-1}}\left[\frac{p_{\tau_i}^G(z)}{p_{\tau_{i-1}}^G(z)} \log \frac{p_{\tau_i}^G(z)}{p_{\tau_{i-1}}^G(z)}\right] p_{\tau_{i-1}}^G(z)dz \\
&= \int_{-\infty}^{\infty} E_{\tau_{i-1}}\left[p_{\tau_i}^G(z) \log p_{\tau_i}^G(z)\right] dz - \int_{-\infty}^{\infty} \log p_{\tau_{i-1}}^G(z) p_{\tau_{i-1}}^G(z)dz \\
&=E_{\tau_{i-1}}\left[\log p_{\tau_i}^G(G_i)\right] - E_{\tau_{i-1}}\left[\log p_{\tau_{i-1}}^G(G_i)\right] \\
&=E_{\tau_{i-1}}\left[\log \frac{p_{\tau_i}^G(G_i)}{p_{\tau_{i-1}}^G(G_i)}\right]
\end{aligned}$$

Using the definition $\mathcal{D}_{KL}(p^G_{\tau_i}(G_i)|p^G_{\tau_{i-1}}(G_i)) = E_{\tau_{i-1}}\left[\log \frac{p^G_{\tau_i}(G_i)}{p^G_{\tau_{i-1}}(G_i)}\right]$ and the law of iterated expectations gives $V^{M,f} = \frac{1}{2}E\left[\int_0^T \left(\theta^{\mathcal{G}}_v\right)^2 dv\right] = \sum_{i=1}^N E\left[\mathcal{D}_{KL}(p^G_{\tau_i}(G_i)|p^G_{\tau_{i-1}}(G_i))\right]$ □

Proof of Proposition 4. The state price densities for the time interval $[\tau_{i-1}, \tau_i]$ are

$$\xi^{\mathcal{G}}_{\tau_{i-1},\tau_i} \equiv \exp\left(-\int_{\tau_{i-1}}^{\tau_i}\left(r_v + \frac{1}{2}\left(\theta^m_v + \theta^{\mathcal{G}}_v\right)^2\right)dv - \int_{\tau_{i-1}}^{\tau_i}\left(\theta^m_v + \theta^{\mathcal{G}}_v\right)dW^{\mathcal{G}}_v\right)$$
$$\xi^m_{\tau_{i-1},\tau_i} \equiv \exp\left(-\int_{\tau_{i-1}}^{\tau_i}\left(r_v + \frac{1}{2}\left(\theta^m_v\right)^2\right)dv - \int_{\tau_{i-1}}^{\tau_i}\theta^m_v dW^m_v\right).$$

where $\theta^{\mathcal{G}}_v = \theta^{\mathcal{G}}_v(G_i)$. Using $dW^{\mathcal{G}}_v = dW^m_v - \theta^{\mathcal{G}}_v dv$ and (A.3) gives $\frac{\xi^{\mathcal{G}}_{\tau_{i-1},\tau_i}}{\xi^m_{\tau_{i-1},\tau_i}} = \frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)}$. It follows that $\frac{\xi^{\mathcal{G}}_T}{\xi^m_T} = \prod_{i=1}^N \frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)}$. Let $E\left[\cdot\,\middle|\,\mathcal{F}_{\tau_{i-1}} \bigvee \sigma(X^b_T)\right] = E_{\tau_{i-1},b_T}[\cdot]$. As

$$E\left[\frac{p^G_{\tau_{n-1}}(G_i)}{p^G_{\tau_n}(G_i)}\,\middle|\,\mathcal{F}_{\tau_n}\bigvee\sigma(X^b_T)\bigvee\sigma(\xi^m_T)\right] = E\left[\frac{p^G_{\tau_{n-1}}(G_i)}{p^G_{\tau_n}(G_i)}\,\middle|\,\mathcal{F}_{\tau_n}\right] = \int_{-\infty}^{\infty}\frac{p^G_{\tau_{n-1}}(z)}{p^G_{\tau_n}(z)}p^G_{\tau_n}(z)dz = 1.$$

The law of iterated expectation gives

$$\begin{aligned}
E_{\tau_0,b_T}\left[\prod_{i=1}^N \frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)}\,\middle|\,\xi^m_T\right] &= E_{\tau_0,b_T}\left[\prod_{i=1}^{N-1}\frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)}E\left[\frac{p^G_{\tau_{N-1}}(G_i)}{p^G_{\tau_N}(G_i)}\,\middle|\,\mathcal{F}_{\tau_N}\right]\,\middle|\,\xi^m_T\right] \\
&= E_{\tau_0,b_T}\left[\prod_{i=1}^{N-2}\frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)}E\left[\frac{p^G_{\tau_{N-2}}(G_i)}{p^G_{\tau_{N-1}}(G_i)}\,\middle|\,\mathcal{F}_{\tau_{N-1}}\right]\,\middle|\,\xi^m_T\right] \\
&\vdots \\
&= E_{\tau_0,b_T}\left[E\left[\frac{p^G_{\tau_0}(G_i)}{p^G_{\tau_1}(G_i)}\,\middle|\,\mathcal{F}_{\tau_1}\right]\,\middle|\,\xi^m_T\right] = 1.
\end{aligned}$$

As $\frac{\xi^{\mathcal{G}}_T}{\xi^m_T} = \prod_{i=1}^N \frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)}$., it follows that

$$E_{\tau_0,b_T}\left[\xi^{\mathcal{G}}_T\,\middle|\,\xi^m_T\right] = \xi^m_T$$
$$E_{\tau_0,b_T}\left[\xi^{\mathcal{G}}_T\right] = E_{\tau_0,b_T}\left[\xi^m_T\right]$$

Let $\epsilon^{\xi} \equiv \xi^m_T\left(\prod_{i=1}^N \frac{p^G_{\tau_{i-1}}(G_i)}{p^G_{\tau_i}(G_i)} - 1\right)$ and note that $\xi^{\mathcal{G}}_T = \xi^m_T + \epsilon^{\xi}$ with $E_{\tau_0,b_T}\left[\epsilon^{\xi}\,\middle|\,\xi^m_T\right]$. Thus ξ^m_T $\mathcal{SSD}$ $\xi^{\mathcal{G}}_T$ in the mean-preserving spread sense. □

Proof of Proposition 5. Suppose the performance fees are of fulcrum type:

$$F(X_T^a, X_T^b) = \alpha X_T^a + \alpha\beta_2 \left(X_T^a - \delta X_T^b\right) \text{ with } \alpha > 0 \text{ and } \beta_2 \geq 0.$$

The definition of certainty equivalent returns gives

$$\begin{aligned} CER^s =& \frac{1}{1-R}\log\left((1-R)E\left[v^f\left(\xi_T^{\mathcal{G}}, X_T^b\right)\right]\right) - \log(X_0^a), \\ CER^u =& \frac{1}{1-R}\log\left((1-R)E\left[v^f\left(\xi_T^{m}, X_T^b\right)\right]\right) - \log(X_0^a). \end{aligned}$$

Thus, the value of the private signals G_i with $i = 1, \ldots, N$ within the period $[0, T]$ to the fund investor with relative risk aversion R is

$$V^f \equiv CER^s - CER^u = \frac{1}{1-R}\log\left(\frac{E\left[v^f\left(\xi_T^{\mathcal{G}}, X_T^b\right)\right]}{E\left[v^f\left(\xi_T^{m}, X_T^b\right)\right]}\right).$$

Let $F^{\xi^m,b}$ (respectively $F^{\xi^{\mathcal{G}},b}$) be the cumulative distribution function (CDF) of ξ_T^m (respectively $\xi_T^{\mathcal{G}}$) based on $\sigma(X_T^b)$, where $\sigma(X_T^b)$ is the filtration generated by the benchmark portfolio at time T. Define $\Delta^{\xi,b}(z) \equiv F^{\xi^{\mathcal{G}},b}(z) - F^{\xi^m,b}(z)$ and $T^{\xi,b}(z) \equiv \int_{-\infty}^{z}\left(F^{\xi^{\mathcal{G}},b}(y) - F^{\xi^m,b}(y)\right)dy$. We have

$$\begin{aligned} \Delta =& E\left[v^f\left(\xi_T^{\mathcal{G}}, X_T^b\right)\right] - E\left[v^f\left(\xi_T^{m}, X_T^b\right)\right] = E\left[E\left[v^f\left(\xi_T^{\mathcal{G}}, X_T^b\right) - v^f\left(\xi_T^{m}, X_T^b\right)\Big|\,\sigma(X_T^b)\right]\right] \\ =& E\left[\int_0^\infty v^f\left(z, X_T^b\right)d\Delta^{\xi,b}(z)\right] = E\left[v^f\left(z, X_T^b\right)\Delta^{\xi,b}(z)\Big|_0^\infty - \int_0^\infty \frac{\partial v^f\left(z, X_T^b\right)}{\partial z}\Delta^{\xi,b}(z)dz\right] \\ =& E\left[-\int_0^\infty \frac{\partial v^f\left(z, X_T^b\right)}{\partial z}\Delta^{\xi,b}(z)dz\right] \\ =& E\left[-\frac{\partial v^f\left(z, X_T^b\right)}{\partial z}T^{\xi,b}(z)\Bigg|_0^\infty + \int_0^\infty \frac{\partial^2 v^f\left(z, X_T^b\right)}{\partial z^2}T^{\xi,b}(z)dz\right] \qquad \text{(A.4)} \\ =& E\left[\int_0^\infty \frac{\partial^2 v^f\left(z, X_T^b\right)}{\partial z^2}T^{\xi,b}(z)dz\right] \qquad \text{(A.5)} \end{aligned}$$

where the last equality follows from the fact $T^{\xi,b}(\infty) = 0$. Together with $T^{\xi,b}(z) \geq 0$ for all $z \in \mathbb{R}_+$, $\Delta = E\left[\int_0^\infty \frac{\partial^2 v^f\left(z, X_T^b\right)}{\partial z^2}T^{\xi,b}(z)dz\right] \leq 0$ (respectively > 0) if $v^f(\cdot, X_T^b)$

is concave (respectively convex). Using $v^f(z, X_T^b) = u(I(y^*z, X_T^b))$, we have

$$\begin{aligned}
&\frac{\partial^2 v^f\left(z, X_T^b\right)}{\partial z^2}\\
=&\frac{du}{dx}(I(y^*z, X_T^b))\frac{\frac{\partial I}{\partial z}(y^*z, X_T^b)^2}{I(y^*z, X_T^b)}\left[\frac{\frac{d^2u}{dx^2}(I(y^*z, X_T^b))I(y^*z, X_T^b)}{\frac{du}{dx}(I(y^*z, X_T^b))}+\frac{\frac{\partial^2 I}{\partial z^2}(y^*z, X_T^b)}{\frac{\partial I}{\partial z}(y^*z, X_T^b)^2}I(y^*z, X_T^b)\right]\\
=&\frac{\frac{\partial I}{\partial z}(y^*z, X_T^b)^2}{I(y^*z, X_T^b)^{R+1}}\left[P^a\frac{\alpha(1+\beta_2)I(y^*z, X_T^b)}{(1-\alpha(1+\beta_2))I^M\left(\frac{y^*z}{\alpha(1+\beta_2)}\right)}-R\right]\\
=&\frac{\left(\frac{1-\alpha(1+\beta_2)}{y^*z^2}\right)^2\left[2\left(1+\frac{\beta_2\delta X_T^b}{1+\beta_2}\frac{y^*z}{1-\alpha(1+\beta_2)}\right)-R\right]}{\left(\frac{1-\alpha(1+\beta_2)}{y^*z}+\frac{\beta_2\delta X_T^b}{1+\beta_2}\right)^{R+1}}
\end{aligned}$$

where $y^* = (1+\beta_2)/X_0^a$ and $P^a \equiv -\frac{d^3u^M}{dx^3}(x)x/\frac{d^2u^M}{dx^2}(x)$ is the relative risk prudence of the manager. The third equality follows because $P^a = 2$ for logarithmic utility and $I(y^*z, X_T^b) = \frac{1-\alpha(1+\beta_2)}{y^*z} + \frac{\beta_2\delta X_T^b}{1+\beta_2}$. Substituting the expression of $\frac{\partial^2 v^f\left(z, X_T^b\right)}{\partial z^2}$ into $(A.5)$ yields

$$\Delta = E\left[\int_0^\infty \frac{\left(\frac{1-\alpha(1+\beta_2)}{y^*z^2}\right)^2\left[2\left(1+\frac{\beta_2\delta X_T^b}{1+\beta_2}\frac{y^*z}{1-\alpha(1+\beta_2)}\right)-R\right]T^{\xi,b}(z)}{\left(\frac{1-\alpha(1+\beta_2)}{y^*z}+\frac{\beta_2\delta X_T^b}{1+\beta_2}\right)^{R+1}}dz\right]$$

□

Proof of Proposition 6. The claims immediately follow from the definition of the value of the private signal. □

Proof of Corollary 1. For $t \in [0, T)$, the conditional density of the signal is

$$p_t^G(x) \equiv \partial_x P_t(G \le x) = \partial_x P_t\left(\log\left(S_{0,T}^m\right) + \sigma^y\int_0^T dW_v^\zeta \le \log(x) + \frac{1}{2}(\sigma^y)^2 T\right).$$

Thus, we have

$$p_t^G(x) = \partial_x P_t\left(\frac{\sigma^m\int_t^T dW_v^m + \sigma^y\int_0^T dW_v^\zeta}{\sqrt{\Sigma_{t,T}}} \le d(x,t)\right) = \partial_x\Phi(d(x,t)) = \frac{\phi(d(x,t))}{x\sqrt{\Sigma_{t,T}}},$$

where

$$d(x,t) = \frac{\log(x) - E_t[\log(G)]}{\sqrt{\Sigma_{t,T}}},$$

$$VAR_t[\log G] = (\sigma^m)^2(T-t) + (\sigma^y)^2 T \equiv \Sigma_{t,T},$$

$$E_t[\log G] = \log S^m_{0,t} + (\sigma^m\theta^m - \frac{1}{2}(\sigma^m)^2)(T-t) - \frac{1}{2}(\sigma^y)^2 T.$$

□

Proof of Corollary 2. As in the proof of Proposition 3

$$\begin{aligned}
&\frac{1}{2}E\left[\int_0^T \left(\theta_v^{\mathcal{G}}\right)^2 dv\right] \\
=&E\left[\log\frac{p_T^G(G_i)}{p_0^G(G_i)}\right] = E\left[\log\left(\frac{\phi(d(G_i,T))}{\phi(d(G_i,0))}\sqrt{\frac{\Sigma_{0,T}}{\Sigma_{T,T}}}\right)\right] = E\left[\log\frac{\phi(d(G_i,T))}{\phi(d(G_i,0))}\right] + \log\sqrt{\frac{\Sigma_{0,T}}{\Sigma_{T,T}}} \\
=&E\left[\frac{d(G_i,0)^2 - d(G_i,T)^2}{2}\right] + \frac{1}{2}\log\left(1+(\sigma^m/\sigma^y)^2\right) = \frac{1}{2}\log\left(1+(\sigma^m/\sigma^y)^2\right).
\end{aligned}$$

Thus, in the presence of fulcrum performance fees, the ex ante value of the private signal G, as described in (3.25), to a fund manager is

$$V^{M,f} \equiv CER^{M,s} - CER^{M,u} = \frac{1}{2}\log\left(1+(\sigma^m/\sigma^y)^2\right).$$

□

Proof of Corollary 3. Using $\xi_T^{\mathcal{G}} = \xi_T^m \frac{p_0^G(G)}{p_T^G(G)}$ and $p_t^G(x) = \frac{\phi(d(x,t))}{x\sqrt{\Sigma_{t,T}}}$, we have

$$
\begin{aligned}
&E\left[\left(\xi_T^{\mathcal{G}}\right)^{R-1}\right] \\
=&E\left[\left(\sqrt{\frac{(\sigma^y)^2}{(\sigma^m)^2+(\sigma^y)^2}}\exp\left(-\left(r+\frac{1}{2}(\theta^m)^2\right)T-\theta^m\sqrt{T}W_1^m+\frac{1}{2}(W_1^\zeta)^2-\frac{(\sigma^m W_1^m+\sigma^y W_1^\zeta)^2}{2\left((\sigma^m)^2+(\sigma^y)^2\right)}\right)\right)^{R-1}\right] \\
=&\left(\frac{(\sigma^y)^2}{(\sigma^m)^2+(\sigma^y)^2}\right)^{\frac{R-1}{2}}\exp\left(-(R-1)\left(r+\frac{1}{2}(\theta^m)^2\right)T+\frac{1}{2}(R-1)^2(\theta^m)^2T\frac{(R-2)(\sigma^m)^2-(\sigma^y)^2}{R(R-2)(\sigma^m)^2-(\sigma^y)^2}\right) \\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}\exp\left(-\frac{1}{2}(w-\mu)^\intercal\Sigma^{-1}(w-\mu)\right)dw^m dw^\zeta \\
=&\left(\frac{(\sigma^y)^2}{(\sigma^m)^2+(\sigma^y)^2}\right)^{\frac{R-1}{2}}\exp\left(-(R-1)\left(r+\frac{1}{2}(\theta^m)^2\right)T+\frac{(R-1)^2(\theta^m)^2T(R-2)(\sigma^m)^2-(\sigma^y)^2}{2\left(R(R-2)(\sigma^m)^2-(\sigma^y)^2\right)}\right)|\Sigma|^{1/2} \\
=&\left(\frac{(\sigma^y)^2}{(\sigma^m)^2+(\sigma^y)^2}\right)^{\frac{R-1}{2}}\sqrt{\frac{(\sigma^m)^2+(\sigma^y)^2}{-R(R-2)(\sigma^m)^2+(\sigma^y)^2}}\exp\left(-(R-1)\left(r+\frac{1}{2}(\theta^m)^2\right)T\right. \\
&\left.+\frac{1}{2}(R-1)^2(\theta^m)^2T\frac{(R-2)(\sigma^m)^2-(\sigma^y)^2}{R(R-2)(\sigma^m)^2-(\sigma^y)^2}\right),
\end{aligned}
$$

where

$$
w=\begin{pmatrix} w^m \\ w^\zeta \end{pmatrix}, \mu=\begin{pmatrix} \theta^m(R-1)\sqrt{T}\frac{(R-2)(\sigma^m)^2-(\sigma^y)^2}{R(R-2)(\sigma^m)^2-(\sigma^y)^2} \\ \theta^m(R-1)\sqrt{T}\frac{(R-1)\sigma^m\sigma^y}{R(R-2)(\sigma^m)^2-(\sigma^y)^2} \end{pmatrix}, \Sigma=\begin{pmatrix} \frac{(R-2)(\sigma^m)^2-(\sigma^y)^2}{R(R-2)(\sigma^m)^2-(\sigma^y)^2} & \frac{-(R-1)\sigma^m\sigma^y}{R(R-2)(\sigma^m)^2-(\sigma^y)^2} \\ \frac{-(R-1)\sigma^m\sigma^y}{R(R-2)(\sigma^m)^2-(\sigma^y)^2} & \frac{-R(\sigma^m)^2-(\sigma^y)^2}{R(R-2)(\sigma^m)^2-(\sigma^y)^2} \end{pmatrix}.
$$

Direct computation yields

$$
E\left[(\xi_T^m)^{R-1}\right]=\exp\left(-(R-1)\left(r+\frac{1}{2}(\theta^m)^2\right)T+\frac{1}{2}(R-1)^2(\theta^m)^2T\right).
$$

Thus, with purely proportional fees $F(X_T^a)=\alpha X_T^a$, the value of a private signal G to investors with relative risk aversion R is

$$
\begin{aligned}
V^p=&\frac{1}{1-R}\log\frac{E\left[\left(\xi_T^{\mathcal{G}}\right)^{R-1}\right]}{E\left[(\xi_T^m)^{R-1}\right]} \\
=&\frac{1}{2}\log\left(1+\left(\frac{\sigma^m}{\sigma^y}\right)^2\right)+\frac{1}{2(R-1)}\log\left(1-\frac{(R-1)^2(\sigma^m)^2}{(\sigma^m)^2+(\sigma^y)^2}\right)+\frac{(R-1)^2(R-2)(\theta^m)^2T(\sigma^m)^2}{2\left(R(R-2)(\sigma^m)^2-(\sigma^y)^2\right)}.
\end{aligned}
$$

According to Proposition 5, $V^p>0$ when $R<2$ and $V^p<0$ when $R>2$. Differentiating V^p with respect to σ^y yields

$$
\frac{\partial V^p}{\partial\sigma^y}=\frac{(R-2)(\sigma^m)^2\left[\left(R(\sigma^m)^2+(\sigma^y)^2\right)\left(1-\frac{(R-1)^2(\sigma^m)^2}{(\sigma^m)^2+(\sigma^y)^2}\right)+(R-1)^2(\theta^m)^2T(\sigma^y)^2\right]}{\sigma^y\left[R(R-2)(\sigma^m)^2-(\sigma^y)^2\right]^2}.
$$

As $s = 1/\sigma^y$, it follows that

$$\frac{\partial V^p}{\partial s} = \frac{(2-R)(\sigma^m)^2\left[(R(\sigma^m)^2+1/s^2)\left(1-\frac{(R-1)^2(\sigma^m)^2}{(\sigma^m)^2+1/s^2}\right)+(R-1)^2(\theta^m)^2T/s^2\right]}{s\left[R(R-2)(\sigma^m)^2-1/s^2\right]^2}.$$

Suppose $R < 1+\sqrt{1+(\sigma^y/\sigma^m)^2}$, which guarantees the ex ante expected utility of a fund investor who delegates his wealth to the skilled manager does not explode, $\frac{\partial V^p}{\partial s} < 0$ for $R > 2$ and $\frac{\partial V^p}{\partial s} > 0$ for $R < 2$. Thus, investors with relative risk aversion $R < 2$ would choose the manager with the highest skill level on the market, and investors with $R > 2$ would prefer the least skilled manager. □

Proof of Corollary 4. The optimal fund value of the uninformed manager at time $t \in [0,T]$ is given by

$$\begin{aligned}
X_t^{a,u^*} =& E\left[\xi_{t,T}^m X_T^{a,s^*} \,\middle|\, \mathcal{F}_t\right] \\
=& \frac{1}{y^{s^*}\xi_t^m} + E\left[\frac{\beta_2\delta}{1+\beta_2}\xi_{t,T}^m X_T^b \mathbb{1}_{\{\xi_T^m<\Psi(X_T^b)\}} \,\middle|\, \mathcal{F}_t\right] + E\left[\frac{\beta_1\delta}{1+\beta_1}\xi_{t,T}^m X_T^b \mathbb{1}_{\{\xi_T^m>\Psi(X_T^b)\}} \,\middle|\, \mathcal{F}_t\right] \\
=& \frac{1}{y^{s^*}\xi_t^m} + \frac{\beta_2\delta}{1+\beta_2}X_t^b E\left[\xi_{t,T}^m X_{t,T}^b \mathbb{1}_{\left\{\xi_{t,T}^m X_{t,T}^b < \frac{\log(\frac{1+\beta_2}{1+\beta 1})}{\xi_t^m X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right\}} \,\middle|\, \mathcal{F}_t\right] \\
&+ \frac{\beta_1\delta}{1+\beta_1}X_t^b E\left[\xi_{t,T}^m X_{t,T}^b \mathbb{1}_{\left\{\xi_{t,T}^m X_{t,T}^b > \frac{\log(\frac{1+\beta_2}{1+\beta 1})}{\xi_t^m X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right\}} \,\middle|\, \mathcal{F}_t\right].
\end{aligned} \quad \text{(A.6)}$$

Since $\int_t^T dW_v^m$ is normally distributed with mean 0 and variance $T-t$ under $\mathcal{F}_t$, replacing $\int_t^T dW_v^m = \sqrt{T-t}z$ yields

$$\begin{aligned}
&\xi_{t,T}^m X_{t,T}^b < \frac{\log(\frac{1+\beta_2}{1+\beta 1})}{\xi_t^m X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)} \\
\Rightarrow \quad & z < \frac{\log\left(\frac{\log(\frac{1+\beta_2}{1+\beta 1})}{\xi_t^m X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right)+\frac{\left(\pi^b\sigma^m-\theta^m\right)^2(T-t)}{2}}{\sqrt{\left(\pi^b\sigma^m-\theta^m\right)^2(T-t)}} \equiv \bar{d}_{1,t}.
\end{aligned}$$

The first expectation on the RHS of (A.6) is

$$E\left[\xi^m_{t,T}X^b_{t,T}\mathbb{1}_{\left\{\xi^m_{t,T}X^b_{t,T}<\frac{\log(\frac{1+\beta_2}{1+\beta1})}{\xi^m_t X^b_t y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right\}}\middle|\mathcal{F}_t\right]$$
$$=\int_{-\infty}^{\bar{d}_{1,t}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\pi^b\sigma^m-\theta^m\right)^2(T-t)+\left(\pi^b\sigma^m-\theta^m\right)\sqrt{T-t}z-\frac{1}{2}z^2\right)=\mathcal{N}\left(d_{1,t}\right),$$

where

$$d_{1,t}=\frac{\log\left(\frac{\log(\frac{1+\beta_2}{1+\beta1})}{\xi^m_t X^b_t y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right)-\frac{1}{2}\left(\pi^b\sigma^m-\theta^m\right)^2(T-t)}{\sqrt{\left(\pi^b\sigma^m-\theta^m\right)^2(T-t)}}.$$

Similar computation applies to the second expectation of (A.6). The Lagrange multiplier y^{u^*} can be obtained by solving $X_0^{a,u^*}=X_0^a$. Plugging the two expectation in (A.6), we obtain

$$X_t^{a,u^*}=\frac{1}{y^{u^*}\xi^m_t}+\frac{\beta_2\delta X^b_t}{1+\beta_2}\mathcal{N}(d_{1,t})+\frac{\beta_1\delta X^b_t}{1+\beta_1}\mathcal{N}(-d_{1,t}).$$

In order to find the optimal fund value of informed manager, we need to find the distribution of W^m_t under private information $\mathcal{G}$. We notice that $W^m_t=W^{\mathcal{G}}_t+\int_0^t\theta^{\mathcal{G}}_v dv$ for $t\in[0,T]$, where

$$\begin{aligned}\theta^{\mathcal{G}}_v&=\sigma^m\left(\frac{\log(G)-E_v[\log G]}{VAR_v[\log G]}\right)\\&=\sigma^m\left(\frac{\log(G)-\left(\sigma^m\theta^m-\frac{1}{2}(\sigma^m)^2-\frac{1}{2}(\sigma^y)^2\right)T-\sigma^m W^m_v}{(\sigma^m)^2(T-t)+(\sigma^y)^2T}\right)\\&=m_vW^m_v+n_v,\end{aligned}$$

where

$$m_v=\frac{1}{v-\left(1+\left(\frac{\sigma^y}{\sigma^m}\right)^2\right)T},n_v=\frac{\left(\theta^m-\frac{(\sigma^m)^2+(\sigma^y)^2}{2\sigma^m}\right)T-\frac{log(G)}{\sigma^m}}{v-\left(1+\left(\frac{\sigma^y}{\sigma^m}\right)^2\right)T}.$$

We have $dW^m_t=(m_tW^m_t+n_t)dt+dW^{\mathcal{G}}_t$. The solution to the stochastic differential equation is

$$W_t^m = \int_0^t \frac{\left(1+\left(\frac{\sigma^y}{\sigma^m}\right)^2\right)T-t}{\left(1+\left(\frac{\sigma^y}{\sigma^m}\right)^2\right)T-v} dW_v^{\mathcal{G}} + \frac{\left(\frac{log(G)}{\sigma^m} - \left(\theta^m - \frac{(\sigma^m)^2+(\sigma^y)^2}{2\sigma^m}\right)T\right)t}{\left(1+\left(\frac{\sigma^y}{\sigma^m}\right)^2\right)T}.$$

This implies that $\int_t^T dW_v^m$ is normally distributed under private information $\mathcal{G}_t$ with mean $\mu_{t,T}$ and variance $\sigma_{t,T}^2$ where

$$\mu_{t,T} = \frac{\sigma^m(T-t)}{\sqrt{\Sigma_{t,T}}} d(G,t),\ \ \sigma_{t,T}^2 = \frac{\Sigma_{T,T}(T-t)}{\Sigma_{t,T}}.$$

The optimal fund value of the informed manager is given by

$$\begin{aligned}
X_t^{a,s^*} =& E\left[\xi_{t,T}^{\mathcal{G}} X_T^{a,s^*} \,\middle|\, \mathcal{G}_t\right] = \frac{1}{y^{s^*}\xi_t^{\mathcal{G}}} + E\left[\frac{\beta_2\delta\xi_{t,T}^{\mathcal{G}}X_T^b}{1+\beta_2}\mathbb{1}_{\left\{\xi_T^{\mathcal{G}}<\Psi^{\mathcal{G}}\right\}} \,\middle|\, \mathcal{G}_t\right] + E\left[\frac{\beta_1\delta\xi_{t,T}^{\mathcal{G}}X_T^b}{1+\beta_1}\mathbb{1}_{\left\{\xi_T^{\mathcal{G}}>\Psi^{\mathcal{G}}\right\}} \,\middle|\, \mathcal{G}_t\right] \\
=& \frac{1}{y^{s^*}\xi_t^{\mathcal{G}}} + \frac{\beta_2\delta}{1+\beta_2} X_t^b E\left[\xi_{t,T}^{\mathcal{G}} X_{t,T}^b \mathbb{1}_{\left\{\xi_{t,T}^{\mathcal{G}}X_{t,T}^b < \frac{\log(\frac{1+\beta_2}{1+\beta_1})}{\xi_t^{\mathcal{G}}X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right\}} \,\middle|\, \mathcal{G}_t\right] \\
&+ \frac{\beta_1\delta}{1+\beta_1} X_t^b E\left[\xi_{t,T}^{\mathcal{G}} X_{t,T}^b \mathbb{1}_{\left\{\xi_{t,T}^{\mathcal{G}}X_{t,T}^b > \frac{\log(\frac{1+\beta_2}{1+\beta_1})}{\xi_t^{\mathcal{G}}X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right\}} \,\middle|\, \mathcal{G}_t\right].
\end{aligned} \tag{A.7}$$

We find that

$$\begin{aligned}
&\xi_{t,T}^{\mathcal{G}} X_{t,T}^b = \xi_{t,T}^m \frac{p_t^{\mathcal{G}}(G)}{p_T^{\mathcal{G}}(G)} X_{t,T}^b \\
=& \sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}} \exp\left(-\frac{1}{2}\left(\pi^b\sigma^m-\theta^m\right)^2(T-t) + \left(\pi^b\sigma^m-\theta^m\right)\int_t^T dW_v^m + \frac{1}{2}\left(d(G,T)^2 - d(G,t)^2\right)\right) \\
=& \sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}} \exp\left(-\frac{\left(\pi^b\sigma^m-\theta^m\right)^2(T-t)}{2} - \frac{\left((\pi^b\sigma^m-\theta^m)\Sigma_{T,T} - \sigma^m\sqrt{\Sigma_{t,T}}d(G,t)\right)^2}{2(\sigma^m)^2\Sigma_{T,T}}\right. \\
&\left. + \frac{d(G,t)^2}{2}\left(\frac{\Sigma_{t,T}}{\Sigma_{T,T}}-1\right) + \frac{(\sigma^m)^2}{2\Sigma_{T,T}}\left(\int_t^T dW_v^m + \frac{(\pi^b\sigma^m-\theta^m)\Sigma_{T,T} - \sigma^m\sqrt{\Sigma_{t,T}}d(G,t)}{(\sigma^m)^2}\right)^2\right) \\
=& \sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}} \exp\left(-\frac{\Sigma_{t,T}}{2}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)^2 + \frac{\left(\sigma^m\int_t^T dW_v^m + \left(\pi^b - \frac{\theta^m}{\sigma^m}\right)\Sigma_{T,T} - \sqrt{\Sigma_{t,T}}d(G,t)\right)^2}{2\Sigma_{T,T}}\right).
\end{aligned}$$

Let $\int_t^T dW_v^m = \mu_{t,T} + \sigma_{t,T} z$ and $\Omega = \frac{\log(\frac{1+\beta_2}{1+\beta 1})}{y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}$, we have

$$\begin{aligned}
&\xi_{t,T}^{\mathcal{G}} X_{t,T}^b < \frac{\Omega}{\xi_t^{\mathcal{G}} X_t^b} \\
&\sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}} \exp\left(-\frac{\Sigma_{t,T}}{2}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)^2 + \frac{\left(\sigma^m \sigma_{t,T} z + \Sigma_{T,T}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)\right)^2}{2\Sigma_{T,T}}\right) < \frac{\Omega}{\xi_t^{\mathcal{G}} X_t^b} \\
&\left(\sigma^m \sigma_{t,T} z + \Sigma_{T,T}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)\right)^2 < \Sigma_{T,T}\left(2\log\left(\sqrt{\frac{\Sigma_{t,T}}{\Sigma_{T,T}}}\Omega\right) + \Sigma_{t,T}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)^2\right) \\
&\bar{d}_{2,t}^- < z < \bar{d}_{2,t}^+,
\end{aligned}$$

where

$$\bar{d}_{2,t}^{\pm} = \frac{\pm\sqrt{\Sigma_{T,T}\left(2\log\left(\frac{\sqrt{\frac{\Sigma_{t,T}}{\Sigma_{T,T}}}\log\frac{1+\beta_2}{1+\beta 1}}{y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right) + \Sigma_{t,T}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)^2\right)} - \Sigma_{T,T}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)}{\sigma^m \sigma_{t,T}}.$$

The first expectation in (A.7) is given by

$$\begin{aligned}
&E\left[\xi_{t,T}^{\mathcal{G}} X_{t,T}^b \mathbb{1}_{\left\{\xi_{t,T}^{\mathcal{G}} X_{t,T}^b < \frac{\log(\frac{1+\beta_2}{1+\beta 1})}{\xi_t^{\mathcal{G}} X_t^b y^{s^*}\delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right\}} \middle| \mathcal{G}_t\right] \\
=&\int_{\bar{d}_{2,t}^-}^{\bar{d}_{2,t}^+} \sqrt{\frac{\Sigma_{T,T}}{2\pi\Sigma_{t,T}}} \exp\left(-\frac{1}{2}\left(\sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}} z - \sqrt{\Sigma_{t,T} - \Sigma_{T,T}}\left(\pi^b - \frac{\theta^m}{\sigma^m} - \frac{d(G,t)}{\sqrt{\Sigma_{t,T}}}\right)\right)^2\right) dz \\
=&\mathcal{N}\left(d_{2,t}^+\right) - \mathcal{N}\left(d_{2,t}^-\right),
\end{aligned}$$

where

$$d_{2,t}^{\pm} = \frac{\pm\sqrt{\Sigma_{T,T}\left(2\log\left(\frac{\sqrt{\frac{\Sigma_{t,T}}{\Sigma_{T,T}}}\log\frac{1+\beta_2}{1+\beta 1}}{y^{s^*}\xi_t^{\mathcal{G}} X_t^b \delta\left(\frac{\beta_2}{1+\beta_2}-\frac{\beta_1}{1+\beta_1}\right)}\right) + \Sigma_{t,T}\left(\pi^b - \frac{\theta^m+\theta_t^{\mathcal{G}}}{\sigma^m}\right)^2\right)} - \Sigma_{t,T}\left(\pi^b - \frac{\theta^m+\theta_t^{\mathcal{G}}}{\sigma^m}\right)}{\sqrt{\Sigma_{t,T} - \Sigma_{T,T}}}.$$

Plugging the expectations in (A.7), we get

$$X_t^{a,s^*} = \frac{1}{y^{s^*}\xi_t^{\mathcal{G}}} + \frac{\beta_2 \delta X_t^b}{1+\beta_2}\left(\mathcal{N}(d_{2,t}^+) - \mathcal{N}(d_{2,t}^-)\right) + \frac{\beta_1 \delta X_t^b}{1+\beta_1}(\mathcal{N}(-d_{2,t}^+) + \mathcal{N}(d_{2,t}^-))$$

The optimal trading strategies of fund managers can be obtained by taking derivatives on both sides of X_t^{a,u^*} (X_t^{a,s^*}) and matching the coefficients in front of dW_t^m. □

Proof of Proposition 7. The first order condition of the fund manager's static problem is

$$((1-\alpha)X_T^{a^*})^{-R^a} = y^*\xi_T^{\mathcal{G}} \Leftrightarrow X_T^{a^*} = \left(y^*\xi_T^{\mathcal{G}}\right)^{-R^a}/(1-\alpha)$$

where y^* is determined by the static budget constraint

$$E\left[\xi_T^{\mathcal{G}}\left(y^*\xi_T^{\mathcal{G}}\right)^{-R^a}\right] = X_0^a.$$

Thus, we have

$$X_T^{a^*} = \frac{X_0^a\left(\xi_T^{\mathcal{G}}\right)^{-1/R^a}}{E_t\left[\left(\xi_T^{\mathcal{G}}\right)^{1-1/R^a}\right]}, F(X_T^{a^*}) = \alpha\left(\xi_T^{\mathcal{G}}\right)^{-1/R^a} X_0^a H_T^{\mathcal{G}}/H_0^{\mathcal{G}}.$$

The manager's optimal fund value at time $t \in [0,T]$ is given by

$$X_t^{a^*} = E_t[\xi_{t,T}^{\mathcal{G}} X_T^{a^*}] = \left(\xi_t^{\mathcal{G}}\right)^{-1/R^a} X_0^a H_t^{\mathcal{G}}/H_0^{\mathcal{G}}$$

where $H_t^{\mathcal{G}} = E_t\left[\left(\xi_{t,T}^{\mathcal{G}}\right)^{1-1/R^a}\right]$.

Using $d(G,t) = \frac{\theta_t^{\mathcal{G}}}{\sigma^m}\sqrt{\Sigma_{t,T}}$ and $d(G,T) = \frac{\theta_t^{\mathcal{G}}\Sigma_{t,T}-(\sigma^m)^2W_{t,T}^m}{\sigma^m\sqrt{\Sigma_{T,T}}}$ we have

$$\begin{aligned}
\xi_{t,T}^{\mathcal{G}} =& \xi_{t,T}^m p_t^G(G)/p_T^G(G) \\
=& \sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}}\exp\left(-\left(r+\frac{1}{2}(\theta^m)^2\right)(T-t) - \theta^m W_{t,T}^m - \frac{d(G,t)^2}{2} + \frac{d(G,T)^2}{2}\right) \\
=& \sqrt{\frac{\Sigma_{T,T}}{\Sigma_{t,T}}}\exp\left(-\left(r+\frac{1}{2}(\theta^m)^2\right)(T-t) - \frac{1}{2}\left(\frac{\theta_t^{\mathcal{G}}}{\sigma^m}\right)^2\Sigma_{t,T} + \frac{1}{2}\left(\frac{\theta_t^{\mathcal{G}}}{\sigma^m}\right)^2\frac{\Sigma_{t,T}^2}{\Sigma_{T,T}}\right. \\
&\left. + \frac{(\sigma^m)^2}{2\Sigma_{T,T}}(W_{t,T}^m)^2 - \left(\frac{\Sigma_{t,T}\theta_t^{\mathcal{G}}}{\Sigma_{T,T}} + \theta^m\right)W_{t,T}^m\right)
\end{aligned}$$

Since $W_{t,T}^m$ is normally distributed with mean 0 and variance $T-t$ under public information, direct computation yields

$$H_t^{\mathcal{G}} = \sqrt{\frac{R^a\Sigma_{T,T}}{\Sigma_{t,T}+(R^a-1)\Sigma_{T,T}}}\left(\frac{\Sigma_{T,T}}{\Sigma_{t,T}}\right)^{-\frac{1}{2R^a}}\exp\left(\left(r+\frac{\Sigma_{t,T}(\theta^m+\theta_t^{\mathcal{G}})^2}{2(\Sigma_{t,T}+(R^a-1)\Sigma_{T,T})}\right)\frac{(1-R^a)(T-t)}{R^a}\right)$$

The manager's optimal trading strategies can be determined by

$$\pi_t^* = \frac{d[X^{a^*}, W^{\mathcal{G}}]_t}{\sigma^m X_t^{a^*} dt} = \frac{\Sigma_{t,T}}{\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T}} \frac{\theta^m + \theta_t^{\mathcal{G}}}{\sigma^m}.$$

Similarly,

$$\pi_t^h = \frac{d[H, W^{\mathcal{G}}]_t}{\sigma^m H_t dt} = \frac{(R^a - 1)(\Sigma_{t,T} - \Sigma_{T,T})}{\Sigma_{t,T} + (R^a - 1)\Sigma_{T,T}} \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m} = \frac{R^a - 1}{1 + \frac{T}{T-t}\left(\frac{\sigma^y}{\sigma^m}\right)^2 R^a} \frac{\theta^m + \theta_t^{\mathcal{G}}}{R^a \sigma^m}.$$

and $\pi_t^m = \pi_t^* - \pi_t^h$. □

Proof of Corollary 5.

$$\begin{aligned}
&E\left[\frac{\left(X_T^{\mathcal{G},a}\right)^{1-R}}{1-R}\right] = E\left[\frac{\left(\left(\xi_T^{\mathcal{G}}\right)^{-\frac{1}{R^a}} x/H_0^{\mathcal{G}}\right)^{1-R}}{1-R}\right] = \frac{x^{1-R}}{1-R} E\left[\left(\left(\xi_T^{\mathcal{G}}\right)^{-\frac{1}{R^a}} /H_0^{\mathcal{G}}\right)^{1-R}\right] \\
=&\frac{x^{1-R}}{1-R} E\left[\left(\sqrt{\frac{R^a(\sigma^y)^2}{(\sigma^m)^2 + R^a(\sigma^y)^2}} \exp\left(-\left(r + \frac{1}{2}(\theta^m)^2\right)T - \frac{1}{R^a}\theta^m \sqrt{T} W_1^m + \frac{1}{2R^a}(W_1^\zeta)^2\right.\right.\right. \\
&\left.\left.\left.-\frac{(\sigma^m W_1^m + \sigma^y W_1^\zeta)^2}{2R^a(\sigma^y)^2} + \frac{\left(\sigma^m \sigma^y W_1^\zeta + (\sigma^m)^2 W_1^m - (R^a - 1)(\sigma^y)^2 \theta^m \sqrt{T}\right)^2}{2R^a(\sigma^y)^2\left((\sigma^m)^2 + R^a(\sigma^y)^2\right)}\right)\right)^{R-1}\right] \\
=&\frac{x^{1-R}}{1-R}\left(\frac{R^a(\sigma^y)^2}{(\sigma^m)^2 + R^a(\sigma^y)^2}\right)^{\frac{R-1}{2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{1}{2}(w-\mu)^\intercal \Sigma^{-1}(w-\mu)\right) dw^m dw^\zeta \\
&\exp\left(-(R-1)(r + \frac{1}{2}(\theta^m)^2)T + \frac{\left((R-1)(R^a + 1 - R)(\sigma^m)^2 + (R + (R^a)^2 - 2R^a)(\sigma^y)^2\right)(R-1)(\theta^m)^2 T}{2\left((1 + R^a - R)R(\sigma^m)^2 + (R^a)^2(\sigma^y)^2\right)}\right) \\
=&\frac{x^{1-R}}{1-R}\left(\frac{R^a(\sigma^y)^2}{(\sigma^m)^2 + R^a(\sigma^y)^2}\right)^{\frac{R-1}{2}} \sqrt{\frac{R^a((\sigma^m)^2 + R^a(\sigma^y)^2)}{R(R^a + 1 - R)(\sigma^m)^2 + (R^a)^2(\sigma^y)^2}} \\
&\exp\left(-(R-1)(r + \frac{1}{2}(\theta^m)^2)T + \frac{\left((R-1)(R^a + 1 - R)(\sigma^m)^2 + (R + (R^a)^2 - 2R^a)(\sigma^y)^2\right)(R-1)(\theta^m)^2 T}{2\left((1 + R^a - R)R(\sigma^m)^2 + (R^a)^2(\sigma^y)^2\right)}\right)
\end{aligned}$$

where $w = (w^m, w^\zeta)^\intercal$ and

$$\mu = \begin{pmatrix} \frac{((R-R^a-1)(\sigma^m)^2 - R^a(\sigma^y)^2)\theta^m(R-1)\sqrt{T}}{R(R-R^a-1)(\sigma^m)^2 - (R^a)^2(\sigma^y)^2} \\ \frac{(R-R^a)\sigma^m\sigma^y\theta^m(R-1)\sqrt{T}}{R(R-R^a-1)(\sigma^m)^2 - (R^a)^2(\sigma^y)^2} \end{pmatrix}, \Sigma = \begin{pmatrix} \frac{-R(\sigma^m)^2 - R^a(\sigma^y)^2}{(\sigma^m)^2 + R^a(\sigma^y)^2} & \frac{(1-R)\sigma^m\sigma^y}{(\sigma^m)^2 + R^a(\sigma^y)^2} \\ \frac{(1-R)\sigma^m\sigma^y}{(\sigma^m)^2 + R^a(\sigma^y)^2} & \frac{(R-R^a-1)(\sigma^m)^2 - (R^a)^2(\sigma^y)^2}{R^a((\sigma^m)^2 + R^a(\sigma^y)^2)} \end{pmatrix}.$$

Direct computation yields

$$
\begin{aligned}
E\left[\frac{(X_T^{m,a})^{1-R}}{1-R}\right] &= E\left[\frac{\left((\xi_T^m)^{-1/R^a}\ x/H_0^m\right)^{1-R}}{1-R}\right] = \frac{x^{1-R}}{1-R}\frac{E\left[(\xi_T^m)^{\frac{R-1}{R^a}}\right]}{(H_0^m)^{1-R}} \\
=&\frac{x^{1-R}}{1-R}\exp\left(-(R-1)(r+\frac{1}{2}(\theta^m)^2)T+\frac{1}{2}\frac{(R+(R^a)^2-2R^a)(R-1)}{(R^a)^2}(\theta^m)^2T\right) \\
=&\frac{x^{1-R}}{1-R}\exp\left(-(R-1)rT+\frac{1}{2}\frac{(R-2R^a)(R-1)}{(R^a)^2}(\theta^m)^2T\right)
\end{aligned}
$$

Thus, with purely proportional fees $F(X_T^a)=\alpha X_T^a$, the value of a private signal G to investors with relative risk aversion R is

$$
\begin{aligned}
V^p =&\frac{1}{2}\log\left(1+\frac{(\sigma^m)^2}{R^a(\sigma^y)^2}\right)+\frac{\log\left(\frac{R^a((\sigma^m)^2+R^a(\sigma^y)^2)}{R(R^a+1-R)(\sigma^m)^2+(R^a)^2(\sigma^y)^2}\right)}{2(1-R)} \\
&+\frac{(R-R^a)^2(R^a+1-R)(\theta^m)^2(\sigma^m)^2T}{2(R^a)^2\left(R(R^a+1-R)(\sigma^m)^2+(R^a)^2(\sigma^y)^2\right)}.
\end{aligned}
$$

$\square$

References

Admati, A. R. and Pfleiderer, P. (1997). Does it all add up? benchmarks and the compensation of active portfolio managers. *Journal of Business*, 70(3):323–350.

Akay, O., Senyuz, Z., and Yoldas, E. (2013). Hedge fund contagion and risk-adjusted returns: A markov-switching dynamic factor approach. *Journal of Empirical Finance*, 22(1):16–29.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

Ang, A. and Bekaert, G. (2004). How do regimes affect asset allocation. *Financial Analysts Journal*, 60(2):86–99.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The crossâĂŘsection of volatility and expected returns. *Journal of Finance*, 61(1):259–299.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Finance*, 91(1):1–23.

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4:313–337.

Aumann, R. J. and Perles, M. (1965). A variational problem arising in economics. *Journal of Mathematical Analysis and Applications*, 25:488–503.

Barber, B. M., Huang, X., and Odean, T. (2016). Which factors matter to investors? evidence from mutual fund flows. *Review of Financial Studies*, 29(10):2600–2642.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55(1):225–264.

Barberis, N. and Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, 68(2):161–199.

Bartram, S. M., Brown, G., and Stulz, R. M. (2016). Why does idiosyncratic risk increase with market risk? NBER Working Paper 22492, National Bureau of Economic Research, Cambridge, MA.

Bekaert, G., Hodrick, R. J., and Zhang, X. (2012). Aggregate idiosyncratic volatility. *Journal of Financial and Quantitative Analysis*, 47(6):1155–1185.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

Berk, J. B. and van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(61):1–20.

Bichuch, M. and Sturm, S. (2014). Portfolio optimization under convex incentive schemes. *Finance Stochastics*, 18(4):873–915.

Billio, M., Sherman, M. G., and Pelizzon, L. (2010). Crises and hedge fund risk. Working Paper, Available at SSRN: https://ssrn.com/abstract=1130742.

Blackburn, D. W., Goetzmann, W. N., and Ukhov, A. D. (2009). Risk aversion and clientele effects. NBER Working Paper 15333, National Bureau of Economic Research, Cambridge, MA.

Brandt, M. W., Brav, A., Graham, J. R., and Kumar, A. (2010). The idiosyncratic volatility puzzle: Time trend or speculative episodes? *Review of Financial Studies*, 23(2):863–899.

Brown, S. J., Goetzmann, W. N., and Ibbotson, R. G. (1999). Offshore hedge funds: Survival and performance, 1989-95. *Journal of Business*, 72(1):91–117.

Campbell, J. Y., Chan, Y. L., and Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics*, 67(1):41–80.

Campbell, J. Y. and Viceira, L. M. (1999). Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics*, 114(2):433–495.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Carpenter, J. N. (2000). Does option compensation increase managerial risk appetite? *Journal of Finance*, 55(5):2311–2331.

Chan, L. K. C., Chen, H.-L., and Lakonishok, J. (2002). On mutual fund investment styles. *Review of Financial Studies*, 15(5):1407–1437.

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105(6):1167–1200.

Cox, J. C. and Huang, C. (1989). Optimum consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49:33–83.

Cox, J. C. and Huang, C. (1991). A variational problem occurring in financial economics. *Journal of Mathematical Economics*, 20:465–487.

Cuoco, D. and Kaniel, R. (2011). Equilibrium prices in the presence of delegated portfolio management. *Journal of Financial Economics*, 101(2):264–296.

Del Guercio, D. and Reuter, J. (2014). Mutual fund performance and the incentive to generate alpha. *Journal of Finance*, 69(4):1673–1704.

Detemple, J. and Rindisbacher, M. (2013). A structural model of dynamic market timing. *Review of Financial Studies*, 26(10):2492–2547.

Elton, E. J., Gruber, M. J., and Blake, C. R. (2012). An examination of mutual fund timing ability using monthly holdings data. *Review of Finance*, 16(3):619–645.

Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, 65(5):1915–1947.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Franco, C. D., Guidolin, M., and Monnier, B. (2017). The robustness of the volatility factor: Linear versus nonlinear factor model. *The Journal of Index Investing*, 8(3):75–88.

Golec, J. and Starks, L. (2004). Performance fee contract change and mutual fund risk. *Journal of Financial Economics*, 732(1):93–118.

Grinblatt, M. and Titman, S. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business*, 62(3):393–416.

Guidolin, M. and Timmermann, A. (2007a). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544.

Guidolin, M. and Timmermann, A. (2007b). Size and value anomalies under regime shifts. *Journal of Financial Econometrics*, 6(1):1–48.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies*, 21(2):889–935.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

Harding, D. and Pagan, A. (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics*, 49(2):365–381.

Harding, D. and Pagan, A. (2006). Synchronization of cycles. *Journal of Econometrics*, 132(1):59–79.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283.

Honda, T. (2003). Optimal portfolio choice for unobservable and regime-switching mean returns. *Journal of Economic Dynamics and Control*, 28(1):45–78.

Hu, Q. and Rindisbacher, M. (2018). A rationale for the clientele effect in money management. Working Paper, Boston University.

Hugonnier, J. and Kaniel, R. (2010). Mutual fund portfolio choice in the presence of dynamic flows. *Mathematical Finance*, 20(2):187–227.

Kacperczyk, M. T., Van Nieuwerburgh, S., and Veldkamp, L. (2008). Time-varying fund manager skill. *Journal of Finance*, 69(4):1455–1484.

Karatzas, I., Lehoczky, J. P., and Shreve, S. E. (1987). Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM Journal on Control and Optimization*, 25(6):1557–1586.

Kim, C.-J. and Nelson, C. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* MIT Press, Cambridge, MA.

Koijen, R. S. (2014). The cross-section of managerial ability and risk preferences. *Journal of Finance*, 69(3):1051–1098.

Li, C. W. and Tiwari, A. (2009). Incentive contracts in delegated portfolio management. *Review of Financial Studies*, 22(11):4681–4713.

Lynch, A. W. (2001). Portfolio choice and equity characteristics: Characterizing the hedging demands induced by return predictability. *Journal of Financial Economics*, 62(1):67–130.

Ma, L., Tang, Y., and Gómez, J.-P. (2016). Portfolio manager compensation and mutual fund performance. unpublished working paper, Northeastern University, University of Florida, IE Business School.

Mamaysky, H., Spiegel, M., and Zhang, H. (2008). Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies*, 21(1):233–264.

Pastor, L. and Stambaugh, R. (2000). Comparing asset pricing models: An investment perspective. *Journal of Financial Economics*, 56(3):335–381.

Pástor, Ľ. and Stambaugh, R. F. (2012). On the size of the active management industry. *Journal of Political Economy*, 120(4):740–781.

Pliska, S. R. (1986). A stochastic calculus model of continuous trading: Optimal portfolios. *Mathematics of Operations Research*, 11(2):371–382.

Russo, F. and Vallois, P. (1993). Forward, backward and symmetric integration. *Probability Theory And Related Fields*, 97(3):403–421.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*, 53(5):1589–1622.

Sotes-Paladino, J. M. and Zapatero, F. (2017). A rationale for benchmarking in money management. unpublished working paper, The University of Melbourne, University of Southern California.

Tu, J. (2010). Is regime switching in stock returns important in portfolio decisions? *Management Science*, 56(7):1198–1215.

Tu, J. and Zhou, G. (2010). Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty. *The Journal of Financial and Quantitative Analysis*, 45(4):959–986.

Turner, C. M., Startz, R., and Nelson, C. R. (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25(1):3–22.

Vial, R. (2015). *Synchronization of Markov Chains in Multivariate Regime-Switching Models*. PhD thesis, University of St. Gallen.

Wachter, J. (2002). Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *Journal of Financial and Quantitative Analysis*, 37(1):63–91.

Whitelaw, R. F. (2000). Stock market risk and return: An equilibrium approach. *Review of Financial Studies*, 13(3):521–547.

CURRICULUM VITAE

[redacted]