

DISCUSSION PAPER SERIES

DP13127
(v. 2)

OPTIMAL FUND MENUS

Julien Hugonnier and Jaksza Cvitanic

FINANCIAL ECONOMICS



OPTIMAL FUND MENUS

Julien Hugonnier and Jaksa Cvitanic

Discussion Paper DP13127
First Published 20 August 2018
This Revision 29 October 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Julien Hugonnier and Jaksa Cvitanic

OPTIMAL FUND MENUS

Abstract

We study the optimal design of a menu of funds by a manager who is required to use linear pricing and does not observe the beliefs of investors regarding one of the risky assets. The optimal menu involves bundling of assets and can be explicitly constructed from the solution to a calculus of variations problem that optimizes over the indirect utility that each type of investor receives. We provide a complete characterization of the optimal menu and show that the need to maintain incentive compatibility leads the manager to offer funds that are inefficiently tilted towards the asset that is not subject to the information friction.

JEL Classification: C62, C71, D42, D82, G11

Keywords: Mutual fund menus, screening, linear pricing, asset bundling

Julien Hugonnier - julien.hugonnier@epfl.ch

Ecole Polytechnique Fédérale de Lausanne, and Swiss Finance Institute and CEPR

Jaksa Cvitanic - cvitanic@caltech.edu

CalTech

Acknowledgements

This work was initiated during a visit of Jaksa Cvitanic to EPFL in the context of the semester on Stochastic Dynamic Models in Mathematical Finance, Econometrics, and Actuarial Sciences organized by the Centre International Bernoulli in the Spring of 2017. We thank Jonathan Berk, Ryan Donnelly, Federico Echenique, Ruediger Fahlenbrach, Damir Filipovic, Benjamin Hebert, Daniel Kuhn, Semyon Malamud, Yaroslav Melnyk, Erwan Morellec, Chris Rogers, and seminar participants at Cambridge, USC and CalTech for conversations on the topic of this paper.

Optimal fund menus*

Jakša Cvitanić[†] Julien Hugonnier[‡]

October 29, 2018

Abstract

We study the optimal design of a menu of funds by a manager who is required to use linear pricing and does not observe the beliefs of investors regarding one of the risky assets. The optimal menu involves bundling of assets and can be explicitly constructed from the solution to a calculus of variations problem that optimizes over the indirect utility that each type of investor receives. We provide a complete characterization of the optimal menu and show that the need to maintain incentive compatibility leads the manager to offer funds that are inefficiently tilted towards the asset that is not subject to the information friction.

Keywords: Mutual fund menus, screening, linear pricing, asset bundling.

JEL Classifications: C62, C71, D42, D82, G11.

*First version: July 2017. This work was initiated during a visit of Jakša Cvitanić to EPFL in the context of the special semester on *Stochastic Dynamic Models in Mathematical Finance, Econometrics, and Actuarial Sciences* organized by the [Centre International Bernoulli](#) in the Spring of 2017. We thank Jonathan Berk, Ryan Donnelly, Federico Echenique, Ruediger Fahlenbrach, Damir Filipovic, Benjamin Hebert, Daniel Kuhn, Semyon Malamud, Yaroslav Melnyk, Erwan Morellec, Chris Rogers, and seminar participants at Cambridge, USC, and CalTech for conversations on the topic of this paper.

[†]Division of the Humanities and Social Sciences, CalTech. M/C 228-77, 1200 East California Blvd., Pasadena CA 91125, USA. Email: cvitanic@hss.caltech.edu.

[‡]EPFL, Swiss Finance Institute, and CEPR. Extranef 212, Quartier UNIL Dorigny, 1015 Lausanne, Switzerland. Email: julien.hugonnier@epfl.ch.

Introduction

One of the most salient characteristics of the mutual fund industry is the proliferation of products that results from the fact that each investment firm offers a large number of funds that often overlap significantly.¹ In this paper, we propose a novel, information based theory of mutual fund families that explains this proliferation. Our reasoning is simple. Consider a firm provides investment services to a population of heterogeneous investors, and assume that the manager of the firm knows the distribution of investors' characteristics but does not observe the type of each individual investor. In such a setting the manager needs to design its offering to screen investors and we claim that, given the constraints to which investment firms are subject to, the optimal strategy is to offer a menu of linearly priced combinations of the risky assets—i.e. funds—constructed to be differentially attractive to different types of investors.

To illustrate this mechanism in the simplest way possible we consider a static model with one riskless asset and two risky assets. The market is populated by a risk neutral investment firm and a continuum of mean-variance investors who can only access the risky assets through the investment firm. Investors agree on the variances of asset returns as well as on the expected return of the first asset but differ in their beliefs regarding the expected return on the second asset. The interpretation of this assumption is that the first risky asset represents a *familiar* investment vehicle such a broad domestic index about which information is freely available and widespread, whereas the second risky asset represents a less familiar asset such as a foreign index about which information is less easily gathered.

The manager of the investment firm knows the distribution of investor beliefs, but does not observe the type of each individual investor. As in a screening model with multiple goods (see, e.g., [Wilson \(1993\)](#) and [Armstrong \(1996\)](#)) the manager is allowed to offer combinations of assets and a corresponding pricing scheme, but we depart from the canonical screening setting in two important ways to take into account the specificities of the mutual fund market. First, we allow investors to combine the funds offered by the manager, subject to a no-short selling constraint. Second, we capture the pervasive use of fraction of fund fees by requiring the manager to use a linear pricing rule that specifies the fees on each fund as a fixed fraction of assets under

¹[Morningstar \(2018\)](#) reports that the 150 largest U.S. fund families jointly offer 7,687 funds of which 4,706 are equity funds, and that the average number of funds offered by a given family in that group is equal to 51 with some sponsors, such as Fidelity, offering more than 400 different funds.

management.² Absent this constraint the well-known screening solution applies and the optimal strategy is to let investors trade the two risky assets separately subject to a linear fee for the familiar asset and a nonlinear pricing scheme with quantity discounts for the non-familiar asset.³ On the contrary, if pricing is required to be linear then the only way for the manager to screen investors is to bundle assets into different funds that each deliver a specific exposure to the underlying risk factors, and one of our main contributions is to show how this optimal fund menu can be constructed.

The solution method we develop consists in three steps. First, we establish that a version of the revelation principle holds in our model. This allows us to restrict the manager to menus in which funds are indexed by investor types and which have the property that, even though he is free to combine funds, each investor finds it optimal to invest only in the fund targeted to his type. In the second step we show that this incentive compatibility constraint can be reduced to a family of differential inequalities and use this formulation to establish that the optimal menu can be characterized in terms of the solution to a constrained calculus of variations problem. In the third and last step we provide a complete analysis of the Euler-Lagrange equations associated with this problem and use the unique solution to these equations to explicitly construct the menu of funds that maximize the total amount of fees.

The analysis of the optimal fund menu allows us to study the combined effect of the information and pricing frictions at play in our model. Consider first the impact of the linear pricing constraint when taking as given the information friction. Absent this constraint the manager optimally offers the two assets separately and grants quantity discount for investment in the non-familiar asset. Imposing linear pricing prevents the manager from using prices to discriminate among investors and instead leads him to rely on bundling as an optimal screening device. Specifically, our results show that the familiar asset is part of the menu because it is not affected by the information friction, but that it is never optimal for the manager to offer the two assets separately. Our findings therefore contribute to the literature on asset bundling (see, e.g., [Adams and Yellen \(1976\)](#), [Spence \(1980\)](#), and [McAfee et al. \(1989\)](#)) by providing conditions under which linear pricing makes mixed bundling optimal.

²The 1970 Amendment to the Investment Advisors Act of 1940 allows managers of mutual funds listed in the U.S. to use performance fees only if they are symmetric around a specified benchmark. As a result, the vast majority of funds in US use linear schedules known as fraction-of-fund fees. For example, [Das and Sundaram \(1998\)](#) report that as of 1998 only 1.4% of funds used performance fees.

³See for example [Mussa and Rosen \(1978\)](#), and [Wilson \(1993\)](#) or [Laffont and Martimort \(2009\)](#) for a textbook treatment. An explicit derivation of the optimal nonlinear pricing scheme in the setting of our mutual fund model is provided in [Appendix A.2](#).

Our characterization of the optimal fund menu can also be used to determine the welfare effect of linear pricing. In particular, we show that this constraint reduces the amount of fees collected by the manager, increases the participation of investors as well as their aggregate welfare, and even results in strict Pareto improvements for all investors if the information friction is not too intense. Therefore, our results provide a justification for regulations, such as the 1970 Amendment to the Investment Advisors Act, that restrict the price setting ability of investment firms.

Given linear pricing and complete information about the beliefs of each investor it is still optimal for the manager to offer a menu of funds. This menu can be derived in closed form and comparing it to that which obtains in the asymmetric information case allows to elicit the effects of the information friction under the linear pricing constraint. In particular, we show that the need to maintain incentive compatibility leads the manager to exhibit bias by proposing funds that are more tilted towards the familiar asset than those he would have offered under complete information. Our theory thus provides an alternative, information-based explanation for two well-documented phenomena in the mutual funds world: closet indexing, that is over-investment in systematic risk, and home bias, that is over-investment in domestic assets.⁴ To provide intuition for this result, we show that, if the full information menu was offered in the asymmetric information case, investors would have an incentive to underreport their beliefs to benefit from the better conditions offered to more pessimistic investors. To prevent this from happening the manager needs to make the funds that target more pessimistic investors less attractive to more optimistic investors, and this is achieved by increasing the share of the familiar asset in all the funds.

Following the usual setting of screening models, our base case model includes a single investment firm that faces a population of investors, as would be the case when considering the provision of retirement accounts to the employees of a given company. To introduce a measure of competition we study an extension of the model in which investors can directly access the familiar asset via an outside fund that charges an exogenously set fee rate. We show that three cases may occur, depending on the level of this exogenous fee rate. If this rate is higher than the optimal fee rate the manager would have offered for the familiar asset absent competition, then the outside fund is dominated and nothing changes. If the outside rate is lower than the optimal rate, but still sufficiently high, then competition leads the manager to exclude a fringe of

⁴See [Cremers and Petajisto \(2009\)](#) [Cremers et al. \(2016\)](#) for background information on the extent and determinants of closet indexing in mutual fund markets, and [Berk and van Binsbergen \(2017\)](#) for an explanation based on decreasing returns to scale.

pessimistic investors from the non-familiar market. The intuition is clear: Pessimistic investors do not care much for the non-familiar asset, at least relative to the familiar one. To entice such investors to allocate part of their capital to a fund that includes both assets, the manager needs to offer conditions that are better than those they can get from the outside fund, and such terms will fail the incentive compatibility constraint. As a result, the manager only offers the familiar asset to these investors and excludes them from the non-familiar market. Despite this exclusion, all investors benefit from the presence of the outside fund because its lower fee rate more than compensates for the lack of exposure to the non-familiar asset. As the fee rate on the familiar asset decreases, investors become less willing to acquire exposure to the familiar asset otherwise than through the outside fund. This makes it gradually harder for the manager to screen investors by bundling, and we show that there is a threshold fee rate below which the optimal strategy is to unbundle the assets. In this case, the optimal menu still excludes a fringe of pessimistic investors but can be implemented by offering the familiar asset at the market fee rate and the non-familiar asset at a constant fee rate. Note, however, that this conclusion is dependent on our specific assumptions. In our model the introduction of the cheap outside fund reduces the optimal number of funds from a continuum down to only two, but this is because there are only two risky assets. If there were more, then it would most likely remain optimal for the manager to offer of a continuum of funds.

In our model investors are not subject to any exclusivity constraint in the sense that they can combine the funds offered by the manager. As part of solving for the optimal menu, we show that this freedom is actually worthless to investors, because the incentive compatibility requires the optimal fund menu to be such that investors never want to combine the funds. One notable exception occurs in the case where the manager faces competition from an outside fund with a sufficiently low fee. In that case the optimal menu can be implemented by offering the assets themselves, and the manager would be better off if he could commit investors to a single fund, because that would allow him to retain the benefits of bundling the assets into funds by preventing investors from mixing.

We emphasize that, while the tractability of our model rests on specific assumptions regarding the preferences and heterogeneity of investors, the main messages of our paper are not dependent on these assumptions. Instead of differing only in their beliefs, investors could well differ along other important dimensions such as risk aversion, initial endowments, the assets they are willing to hold, or the risks they

are exposed to. Furthermore, preferences need not be quadratic and there may exist more than two risky assets. Such generalizations are likely to make the model a lot more difficult to solve because they lead to multidimensional screening problems (see, e.g., [Rochet and Choné \(1998\)](#)), but they would not undermine the validity of our mechanism. In particular, the linear pricing constraint would still lead the manager to use asset bundling as an optimal screening device, and the need to maintain incentive compatibility would still generate a loss of efficiency.

Our paper relates to a large theoretical literature on mutual funds and delegated portfolio management. [Hugonnier and Kaniel \(2010\)](#) study a model close to ours but in which the fund manager faces a single investor about whom he has full information. [Breton et al. \(2010\)](#) extend the model of [Hugonnier and Kaniel \(2010\)](#) to the case where two managers compete and show that competition does not benefit investors because, in equilibrium, the funds offered by the two managers are colinear. In our model we take as given the fact that the pricing of funds must be linear. By contrast, [Admati and Pfleiderer \(1997\)](#), [Carpenter \(2000\)](#), [Das and Sundaram \(2002\)](#), [Basak et al. \(2007\)](#), [Cuoco and Kaniel \(2011\)](#), and [Basak and Pavlova \(2013\)](#) study the effects of different exogenous fee structures on allocations, social welfare, risk-taking, market efficiency, and asset prices, while [Bhattacharya and Pfleiderer \(1985\)](#), [Ou-Yang \(2003\)](#), [Dybvig et al. \(2010\)](#), and [Cvitanić and Xing \(2018\)](#) among others, adopt an optimal contracting perspective in which investors control the compensation of the fund manager. Our model focuses on the design of an optimal fund menu in a static setting where the customer base is fixed. Therefore, it abstracts from some important dynamic considerations such as learning about managerial skill and its implications for the relation between past performance and fund flows. Examples of papers that examine the impact of investors' learning about managerial skills and/or technology include [Lynch and Musto \(2003\)](#), [Dangl et al. \(2008\)](#), [Carlin and Manso \(2011\)](#), [Pastor and Stambaugh \(2012\)](#), and [Brown and Wu \(2016\)](#) among others.

There are a few papers that explicitly model fund families. [Mamaysky and Spiegel \(2002\)](#) propose a model that explains the existence of many different fund families, while we focus on why there are many funds inside one family. In their model, each investment company gathers information that is specific and offers portfolios aimed at the subset of the population to which that information is most useful. In our model, the funds inside a given family adapt to the beliefs among its population of investors, but our findings are not necessarily at odds with those of [Mamaysky and Spiegel \(2002\)](#). In particular, they empirically document that when an investment firm introduces a

new fund, it typically uses a strategy that places this fund in a different Morningstar category than its existing ones, which is in line with the fact that in our model that a new fund would only be introduced following a change in the customer base of the family. Our findings are also in agreement with [Gruber \(1996\)](#), [Khorana and Servaes \(2012\)](#), and [Massa \(2000\)](#) who show both empirically and theoretically that product differentiation is an effective strategy for investment firms to maximize revenues. In recent work [Brown and Wu \(2016\)](#) follows the approach of [Berk and Green \(2004\)](#) to develop a continuous-time model in which the performance of the funds offered by a sponsor carries information about the common skills and resources shared across the whole family, while [Berk et al. \(2017\)](#) propose a model of an investment firm that allocates its investors' capital to a population of heterogeneous fund managers who can each add value to the firm subject to decreasing returns to scale.

Our paper also contributes to the literature on screening and asset bundling, see for example [Adams and Yellen \(1976\)](#), [Spence \(1980\)](#), [McAfee et al. \(1989\)](#), [Wilson \(1993\)](#), [Armstrong \(1996\)](#), and [Stole \(2001\)](#) among others. In particular, our paper can be seen as multiple goods extension of the model of [Mussa and Rosen \(1978\)](#) in which the monopolist is required to use linear pricing. Because of this constraint, the monopolist cannot resort to nonlinear pricing as a mean of discriminating among his customers. Instead, she will use product bundling and our contribution to this literature is to show how the optimal menu of linearly priced bundles can be constructed. To the best of our knowledge, this paper is the first to analytically derive a solution to a screening problem with multiple goods and a linear pricing constraint. In a related contribution [Rothschild \(2015\)](#) also considers a screening problem with linear pricing, but his graphical analysis is limited to qualitative properties of the optimum.

The remainder of the paper is organized as follows. In Section 1 we present the model. In Section 2 we show how the design of an optimal fund menu can be reduced to the solution of a calculus of variations problem, provide a complete description of the optimal fund menu and analyze its most salient properties. Finally, in Section 3 we extend the base case model to allow investors a direct access to an outside fund that offers the familiar asset. Section 4 concludes. Appendix A derives the solution to our model in three important benchmark cases: the frictionless case where investors can freely access all assets, the asymmetric information case where the manager is allowed to use any pricing scheme, and the full information case where he is required to use linear pricing. Appendices B and C gather all the proofs.

1 The model

We consider a static model of a financial market that consists in three assets: A riskless asset with gross return r and two risky assets whose gross *excess* returns are given by a vector $\epsilon \in \mathbf{R}^2$ of independent variables with unit variances. We interpret the first risky asset as representing a widely familiar asset, such as a broad domestic market index, about which investors have homogenous beliefs and the second one as being a less familiar asset about which investors have heterogenous beliefs.

The market is populated by a single risk-neutral fund manager and a unit measure of risk-averse investors. All market participants agree that returns have unit variances and that the expected gross excess return on the familiar asset, or equivalently its risk premium, is given by ζ for some constant $\zeta > 0$, but investors differ in their beliefs regarding the expected return on the other asset. Specifically, we assume that each investor is associated with a type $\theta \in \Theta := [0, \theta_H]$ that represents her perception of the expected gross excess return on the non-familiar asset. Each investor knows her own type but the only information available to the manager is that types are uniformly distributed over Θ . Investors have initial wealth $w_0 \equiv 1$ and mean-variance preferences over terminal wealth. Specifically, we assume that the utility that an investor of type $\theta \in \Theta$ derives from terminal wealth w_1 is given by

$$u(\theta, w_1) := aE_{\theta}[w_1 - rw_0] - \frac{a^2}{2}\text{var}_{\theta}[w_1]$$

where $a > 0$ captures the investors' risk-aversion, and the subscript indicates that the computation of the mean and variance is performed under the probability measure P_{θ} associated with the investor's belief regarding the non-familiar asset.

Investors can freely trade the riskless asset but can only access the risky assets through the fund(s) offered by the manager.⁵ Within a given class of mutual fund shares, fees are typically specified as a constant percentage of the initial value of assets under management. We capture this important feature by assuming that the fund manager is only allowed to use *linear price schedules* that charge investors a constant fraction of the initial investment. Accordingly, a fund is specified by a pair (γ, ϕ) where $\gamma \in \mathbf{R}_+$ is the fee that the manager collects at the terminal time per dollar

⁵The fact that investors have to go through the manager to access the risky assets may be due to transaction and/or informational costs. For example, one may think that investors form their beliefs about the non-familiar asset using information provided as part of the onboarding process and that this information would not be otherwise available to investors.

invested in the fund, and the vector $\phi \in \mathbf{R}^2$ represents the amounts invested in the two risky assets per dollar of assets under management. When offered by the manager, each such pair gives rise to a new risky asset that investors can allocate capital to and which provides a gross excess return given by

$$R(\gamma, \phi) := \phi^\top \varepsilon - \gamma.$$

where the superscript \top denotes transposition. Since it is generally not possible to short a mutual fund, we assume that investors are subject to a trading constraint that prevents them from taking short positions in any of the offered funds.⁶

Because the pricing of funds is constrained to be linear, the manager cannot rely on quantity discounts to screen investors as he would in the standard model of monopoly pricing under asymmetric information.⁷ Instead, he will exploit the fact that investors have different preferences for the risky assets by offering a menu of linearly priced funds that represent different combinations of exposures to these assets.

Definition 1 A fund menu is a collection $\mathbf{m} = (\gamma, \phi, \mathcal{M})$ where \mathcal{M} is a set that indexes the funds in the menu and $(\gamma, \phi) : \mathcal{M} \rightarrow \mathbf{R}_+ \times \mathbf{R}^2$ are functions that represent the fee rate and the vector of loadings of the funds on the two risky assets.

Let \mathbf{m} be a given fund menu. In addition to linear pricing, another key feature of the mutual fund market compared to a standard screening environment is that investors are not constrained to pick a single fund and can in fact combine different funds to achieve their preferred exposure. We capture this *non exclusivity* by taking as the action set of investors the space $\mu_+(\mathcal{M})$ of nonnegative measures on \mathcal{M} and associate to each element q of this set the terminal wealth given by

$$w_1(q, \mathbf{m}) := rw_0 + \int_{\mathcal{M}} R(\gamma(m), \phi(m)) q(dm).$$

The optimization problem of an investor of type $\theta \in \Theta$ who takes the fund menu \mathbf{m} as

⁶We do not require that ϕ_1 and ϕ_2 be positive or that they sum up to one. As a result, the offered funds may include short risky asset positions as well as investments in the riskless asset. The latter is an additional, indirect source of inefficiency because it implies that investors may end up paying the manager for an investment that they could have done on their own. However, and as we will see in Section 2.3.2 below, this indirect inefficiency does not materialize because there always exist an *all-equity* implementation of the optimal menu in which none of the offered funds invest in the riskless asset.

⁷See for example [Mussa and Rosen \(1978\)](#) and [Laffont and Martimort \(2009\)](#) for a textbook treatment. A derivation of the optimal nonlinear pricing scheme in the setting of our delegated portfolio management model is provided in [Appendix A.2](#).

given is then defined by

$$v_i(\theta, \mathbf{m}) := \sup_{q \in \mu_+(\mathcal{M})} u(\theta, w_1(q, \mathbf{m})) \quad (1)$$

and the aggregation of the investors' individual portfolio decisions generates a total amount of management fees given by

$$v_m(\mathbf{m}) := (1/\theta_H) \int_{\Theta \times \mathcal{M}} \gamma(m) q^*(dm; \theta, \mathbf{m}) d\theta \quad (2)$$

where the measure $q^*(\theta, \mathbf{m})$ is the best response of an investor of type $\theta \in \Theta$ to the menu \mathbf{m} . In accordance with the above definitions, a menu \mathbf{m}^* is said to be optimal if it maximizes the total amount of fees in (2).

Remark 1 The assumption that returns are independent and have unit variance simplifies the presentation, but does not entail any loss in generality. Indeed, a direct calculation shows that investing through a fund (γ, ψ) in two correlated risky assets with non unit variances is equivalent to investing in two uncorrelated risky assets with unit variances through the modified fund defined by

$$(\gamma, \phi) = \left(\gamma, \begin{bmatrix} \sigma_1 & \rho\sigma_1 \\ 0 & \sigma_2\sqrt{1-\rho^2} \end{bmatrix} \psi \right)$$

where the constants $\sigma_1, \sigma_2 > 0$ and $\rho \in [-1, 1]$ denote, respectively, the standard deviations of asset returns and their correlation coefficient.

Remark 2 Our results remain unchanged if we consider a continuous-time market with iid normal returns and CARA investors who do not update their beliefs regarding the non-familiar asset and assume that the manager only offers funds whose composition is non-random. The assumption that investors do not learn seems somewhat stark, but may be justified if the proportion of investors with the same beliefs remains constant even if the investors do change beliefs at the individual level, or if the manager keeps changing the features of the funds she offers to roll back the learning of the investors, as in the optimal obfuscation model of [Carlin and Manso \(2011\)](#).

2 Solution

2.1 The revelation principle

To narrow down the search for the optimal menu we start by showing that, despite the linear pricing and non-exclusivity of funds, a version of the revelation principle holds in our setting. Let $\zeta(\theta) := (\zeta, \theta)^\top$ be the vector of expected gross excess returns (i.e. risk premia) perceived by an investor of type $\theta \in \Theta$ and denote by

$$\pi(\theta, \phi) := \arg \max_{q \in \mathbf{R}_+} u(\theta, rw_0 + qR(1, \phi)) = \frac{1}{a\|\phi\|^2} \left(\phi^\top \zeta(\theta) - 1 \right)_+$$

with $x_+ := \max\{0, x\}$, the amount of capital that such an investor would optimally invest in the fund with characteristics $(1, \phi)$ when allocating his wealth only between the riskless asset and that fund. Our first result shows that the manager can without loss of generality restrict her attention to the set of menus such that funds are indexed by investors types, fee rates are normalized to one, and each investor finds it optimal to only invest in the fund targeted to his type.

Proposition 1 *Given any menu $\bar{\mathbf{m}}$, there exists a menu $\mathbf{m} = (\gamma, \phi, \mathcal{M})$ such that:*

1. $\mathcal{M} = \Theta$;
2. $\gamma(\theta) = 1$ for all $\theta \in \Theta$;
3. $q^*(\sigma; \theta, \mathbf{m}) = \mathbf{1}_{\{\theta \in \sigma\}} \pi(\theta, \phi(\theta))$ for all $\theta \in \Theta$ and $\sigma \subseteq \Theta$.
4. $v_m(\mathbf{m}) = v_m(\bar{\mathbf{m}})$;
5. $v_i(\theta, \mathbf{m}) = v_i(x, \bar{\mathbf{m}})$ for all $\theta \in \Theta$.

Properties 4 and 5 mean that the manager and the investors are indifferent between the original menu and the new menu which satisfies properties 1 to 3. Normalizing the fee rate serves two purposes. First, it reduces the choice of the manager to that of a *fund loading function* $\phi : \Theta \rightarrow \mathbf{R}^2$. Second, it allows to easily compare the funds in a given menu: when all funds share the same fee rate, they differ only in the risk profile that they offer. Property 3 requires that each investor finds it optimal to invest only in the riskless asset and the fund $(1, \phi(\theta))$ that is specifically aimed at his type. This condition is the analogue in our setting of the *incentive compatibility constraint* in classical screening problems.

In view of the proposition, the manager's optimization problem can be reduced to the maximization of the integral

$$I(\phi) := (1/\theta_H) \int_{\Theta} \pi(\theta, \phi(\theta)) d\theta \quad (3)$$

over the set Φ_0 of fund loading functions that are incentive compatible. Our next result provides a variational characterization of incentive compatibility that will be instrumental in our construction of the optimal fund menu.

Proposition 2 *A fund loading function $\phi : \Theta \rightarrow \mathbf{R}^2$ is incentive compatible if and only if*

$$\phi(\theta')^\top \xi(\theta) - 1 - \frac{\phi(\theta)^\top \phi(\theta')}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \xi(\theta) - 1 \right)_+ \leq 0 \quad (4)$$

for all types $(\theta, \theta') \in \Theta^2$.

To understand the above result assume that the manager offers a loading function $\phi : \Theta \rightarrow \mathbf{R}^2$. If an investor of type $\theta \in \Theta$ perceives that the risk premium

$$E_\theta [R(1, \phi(\theta))] = \phi(\theta)^\top \xi(\theta) - 1$$

on the fund targeted to him is nonpositive, then (4) requires that this investor also perceives all the other funds in the menu as offering negative risk premia and, thus, finds it optimal to only invest in the riskless asset. On the other hand, if the investor perceives that the risk premium on the fund targeted to him is strictly positive, then (4) requires that, for any $\theta' \in \Theta$, the alpha

$$\begin{aligned} \alpha(\theta, \theta') &:= E_\theta [R(1, \phi(\theta'))] - \frac{\text{cov}_\theta [R(1, \phi(\theta)), R(1, \phi(\theta'))]}{\text{var}_\theta [R(1, \phi(\theta))]} E_\theta [R(1, \phi(\theta))] \\ &= \left(\phi(\theta')^\top \xi(\theta) - 1 \right) - \frac{\phi(\theta)^\top \phi(\theta')}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \xi(\theta) - 1 \right) \end{aligned}$$

of fund $(1, \phi(\theta'))$ relative to fund $(1, \phi(\theta))$ be negative, so that including the other fund in his portfolio does not increase risk-adjusted performance.

2.1.1 First best with linear pricing

To understand the implications of the information friction to which the manager is subject, it is useful to briefly consider the *first best* case in which the manager knows the type of each investor, but is still required to use linear pricing. In this case, we

show in Appendix A.3 that it is optimal for the manager to offer investors of type $\theta \in \Theta$ a single fund with loading function

$$\phi^\circ(\theta) = \frac{2\zeta(\theta)}{\|\zeta(\theta)\|^2}$$

and fee rate equal to one. Substituting this fund loading function into the incentive compatibility condition (4) shows that

$$\phi^\circ(\theta')^\top \zeta(\theta) - 1 - \frac{\phi^\circ(\theta)^\top \phi^\circ(\theta')}{\|\phi^\circ(\theta)\|^2} \left(\phi^\circ(\theta)^\top \zeta(\theta) - 1 \right) = \frac{\theta'(\theta - \theta')}{\|\zeta(\theta')\|^2}$$

is nonnegative for any pair of types $(\theta, \theta') \in \Theta^2$ such that $\theta' \leq \theta$. This shows that the first best fund menu is not incentive compatible and reveals that the adverse selection problem facing the manager is that, when offered the first best menu, any given investor has an incentive to pretend to be less optimistic than he really is in order to benefit from better investment conditions.

2.1.2 The exclusive case

Consider now the *exclusive* case in which each investor can allocate capital to at most one fund. Given a menu satisfying properties 1 and 2 of Proposition 1, the optimal strategy of an investor of type $\theta \in \Theta$ who allocates his wealth between fund $(1, \phi(\theta'))$ and the riskless asset is given by

$$\arg \max_{q \in \mathbf{R}_+} u(\theta, rw_0 + qR(1, \phi(\theta'))) = \pi(\theta, \phi(\theta'))$$

and delivers him the indirect utility

$$v(\theta, \theta') := u(\theta, rw_0 + \pi(\theta, \phi(\theta'))R(1, \phi(\theta'))) = \frac{1}{2} \left(\frac{\phi(\theta')^\top \zeta(\theta) - 1}{\|\phi(\theta')\|} \right)_+^2. \quad (5)$$

Under exclusivity, incentive compatibility requires that each investor finds it optimal to pick the fund targeted to him in the sense that

$$v(\theta) := v(\theta, \theta) = \sup_{\theta' \in \Theta} v(\theta, \theta'), \quad \theta \in \Theta. \quad (6)$$

The Cauchy-Schwartz inequality implies that this incentive compatibility condition is weaker than its non exclusive counterpart in (4) so that, as would be expected, the manager cannot do worse when the investors are forced to commit to a single fund. However, since

$$\left. \frac{dv(\theta, \theta')}{d\theta'} \right|_{\theta'=\theta} = \pi(\theta, \phi(\theta)) \left. \frac{d\alpha(\theta, \theta')}{d\theta'} \right|_{\theta'=\theta}$$

we have that the first order conditions induced by those two constraints coincide and we will see below that the same menu is optimal under either constraint. This result does not mean that the ability to commit investors to a single fund is always worthless to the manager. In particular, we show in Section 3 that the exclusive and non exclusive solutions no longer coincide if investors can directly trade in the familiar asset at a sufficiently low fee.

2.2 The relaxed problem

Propositions 1 and 2 imply that the manager's problem reduces to the maximization of (3) over the set Φ_0 of fund loading functions that satisfy (4). To solve that problem we further restrict the manager's choice set by imposing the technical requirement that the fund loading function belongs to the intersection $\Phi := \Phi_0 \cap AC(\Theta; \mathbf{R}^2)$ of Φ_0 with the space of absolutely continuous functions on Θ with values in \mathbf{R}^2 . The optimization problem that we consider is therefore given by

$$M := \sup_{\phi \in \Phi} I(\phi). \tag{P}$$

The main difficulty in solving this problem arises from the fact that the condition (4) which defines the feasible set cannot be dealt with using standard techniques because it involves the values of the unknown vector-valued function at all points of the type space. To overcome this difficulty, we follow the *first order approach* (see, e.g., [Mirrlees \(1971\)](#), [Rochet \(1987\)](#), and [Rochet and Choné \(1998\)](#)), which exploits the first order condition induced by the incentive compatibility constraint to show that, instead of optimizing over fund loading functions, the manager can optimize over the indirect utility $v(\theta)$ and marginal utility $\dot{v}(\theta)$ that her menu of funds delivers to each type of investor. Our first result in this direction relates incentive compatible fund loading functions to the indirect utility functions they induce.

Lemma 1 Assume that $\phi \in \Phi$ is incentive compatible. Then the indirect utility function defined by (5) and (6) belongs to the space $AC(\Theta; \mathbf{R})$. Furthermore, it satisfies

$$2v(\theta) \geq [\dot{v}(\theta)]^2, \quad (7)$$

and the corresponding optimal investment can be expressed as

$$\pi(\theta, \phi(\theta)) = (1/a) \left(\theta \dot{v}(\theta) - 2v(\theta) + \xi \sqrt{2v(\theta) - [\dot{v}(\theta)]^2} \right), \quad (8)$$

for almost every type $\theta \in \Theta$.

Let $\phi \in \Phi$ be an incentive compatible fund loading function. Relying on the above lemma we have that the total amount of fees generated by the investors' best responses to the corresponding menu is given by

$$\theta_H I(\phi) = \int_{\Theta} \pi(\theta, \phi(\theta)) d\theta = \int_{\Theta} F(\theta, v(\theta), \dot{v}(\theta)) d\theta$$

where

$$F(\theta, v(\theta), \dot{v}(\theta)) := (1/a) \left(\theta \dot{v}(\theta) - 2v(\theta) + \xi \sqrt{2v(\theta) - [\dot{v}(\theta)]^2} \right).$$

and v is the indirect utility function associated to ϕ through (8). It follows that the manager's value function satisfies

$$\theta_H M = \sup_{\phi \in \Phi} (\theta_H I(\phi)) \leq V := \sup_{v \in \mathcal{V}} \int_{\Theta} F(\theta, v(\theta), \dot{v}(\theta)) d\theta \quad (\mathcal{R})$$

where \mathcal{V} denotes the set of functions $v \in AC(\Theta; \mathbf{R})$ that satisfy (7). Following the usual terminology of screening models, we refer to (\mathcal{R}) as the *relaxed problem* because it only takes into account the first order condition induced by the incentive compatibility constraint. Our goal will be to show that at the optimum of the relaxed problem this first order condition is sufficient for incentive compatibility so that the solution to (\mathcal{P}) can be constructed from the solution to (\mathcal{R}).

Compared to the original optimization problem (\mathcal{P}), the relaxed problem (\mathcal{R}) is a scalar calculus of variations problem that can be solved using standard techniques (see, e.g., [Mesterton-Gibbons \(2009\)](#)). Specifically, using subscripts to denote partial derivatives, we have that a necessary condition for optimality is given by the Euler

Lagrange equation

$$F_{v(\theta)}(\theta, v(\theta), \dot{v}(\theta)) - \frac{d}{d\theta} F_{\dot{v}(\theta)}(\theta, v(\theta), \dot{v}(\theta)) = 0, \quad \theta \in \Theta, \quad (9)$$

and, because the values of the indirect utility function $v(\theta)$ on the boundaries of the type space are free, this second order differential equation should be solved subject to the boundary conditions

$$F_{\dot{v}(\theta)}(\theta, v(\theta), \dot{v}(\theta)) = 0, \quad \theta \in \{0, \theta_H\}.$$

Calculating the derivatives involved in these expressions, and simplifying the result leads to the boundary value problem

$$v(\theta)(1 + \ddot{v}(\theta)) - [\dot{v}(\theta)]^2 = \frac{3}{2\zeta} \left(2v(\theta) - [\dot{v}(\theta)]^2 \right)^{\frac{3}{2}}, \quad \theta \in \Theta \quad (10)$$

subject to

$$\dot{v}(0) = 0, \quad (11)$$

$$\theta_H = \zeta \dot{v}(\theta_H) \left(2v(\theta_H) - [\dot{v}(\theta_H)]^2 \right)^{-\frac{1}{2}}. \quad (12)$$

Our next result establishes existence, uniqueness, and some properties of the solution to this boundary value problem and verifies that this solution attains the supremum in the relaxed problem.

Proposition 3 *There exists a unique solution $v^* \in C^2(\Theta; \mathbf{R})$ to the boundary value problem defined by (10), (11) and (12). This solution is strictly increasing, strictly convex, and attains the supremum in the relaxed problem.*

2.3 The optimal fund menu

By definition we have that the value of the relaxed problem (\mathcal{R}) gives an upper bound on the value of the manager's problem (\mathcal{P}). To show that these values coincide, and thus arrive at a complete characterization of the optimal menu, we need to construct a fund loading function $\phi^* \in \Phi$ that delivers to each investor the indirect utility prescribed for his type by the solution to the relaxed problem. The following theorem provides such a construction and constitutes one of our main results.

Theorem 1 Denote by $v^* \in C^2(\Theta; \mathbf{R})$ the solution to the relaxed problem (\mathcal{R}) as defined in Proposition 3. Then,

$$F^*(\theta) := F(\theta, v^*(\theta), \dot{v}^*(\theta)) > 0, \quad \theta \in \Theta, \quad (13)$$

and the fund loading function defined by

$$\phi^*(\theta) := \frac{1}{aF(\theta, v^*(\theta), \dot{v}^*(\theta))} \left(\sqrt{2v^*(\theta) - \dot{v}^*(\theta)^2}, \dot{v}^*(\theta) \right)^\top$$

attains the supremum in (\mathcal{P}) . In particular, every investor type invests a strictly positive amount in the fund targeted to his type.

Our next result provides basic comparative statics for the optimal fund menu and the amount of fees that the manager receives from each type of investor.

Proposition 4 The function $\phi_1^*(\theta)$ is decreasing in θ while the functions $F^*(\theta)$ and $\phi_2^*(\theta)$ are increasing.

The above proposition shows that the offered fund become gradually become more tilted towards the non-familiar asset as θ increases and that the manager receives more fees from more optimistic investors. These results are intuitive. Indeed, investors with higher θ are more interested in the non-familiar asset. Knowing this, the manager gradually tilts the exposure of his funds towards the non-familiar risk to extract more fees from more optimistic investors. To guarantee that investors do not have any incentive to deviate from the fund targeted to them, the manager needs to construct his optimal menu in such way as to provide more optimistic investors with a disproportionately larger indirect utility than less optimistic investors. This is why the indirect utility function $v^*(\theta)$ induced by the optimal menu is not only strictly increasing, but also strictly convex in the investor's type.

The comparative statics of the various equilibrium outcomes with respect to the risk premium ξ of the familiar asset and the range θ_H of perceived risk premia on the non-familiar asset are a lot more difficult to derive analytically. As shown by the following lemma, a notable exception concerns the manager's welfare.

Proposition 5 The manager always prefers to face more optimistic investors in the sense that his value function M is increasing in both ξ and θ_H .

The mechanism behind the above result is clear: As ζ or θ_H increase investors become more eager to invest in the risky assets and thus are willing to pay the manager a larger amount to get access to those assets through the funds.

The right panel of Figure 1 shows that the indirect utility of investors depends positively on the risk premium ζ of the familiar asset and negatively on the range θ_H of perceived risk premia on the non-familiar asset. The first result is partly due to the fact that an increase in ζ implies a reduction in the relative importance of the information friction that affects the non-familiar asset and, thus, leads to an increase in the welfare of investors. To understand the second result recall that, in order to satisfy the incentive compatibility constraint, the manager has to make sure that no investor has an incentive to under-report his type by switching to a fund that is targeted to less optimistic investors. Now fix an arbitrary type $\theta \in \Theta$. As the upper bound of the type space increases, the investors whose type are larger than θ and who have to be deterred from under-reporting their type as θ , become more optimistic on average. To prevent these investors from under-reporting the manager then needs to modify the menu to worsen the conditions he offers to investors of type θ and this explains why investors benefit from being part of a narrower customer base.

The left panel of Figure 1 illustrates the effect of a change in the range θ_H of perceived risk premia on the indirect utility of an investor who stands at a given percentile in the distribution. As shown by the right panel of the figure, an increase in θ_H leads to a decrease in the utility of all investors. However, another effect comes into play when considering an investor at a given percentile because, as θ_H increases, the type and hence the indirect utility of investors at a given percentile also increase. As shown by the left panel of the figure, the first effect dominates at low percentiles while the second dominates at higher percentiles.

To elicit the impact of the information friction on the optimal menu and the welfare of the various players, we now compare the outcomes of the model to those that would prevail in the first best benchmark introduced in Section 2.1.1. The results in Appendix A.3 show that in this case the optimal fund loading $\phi^\circ(\theta)$, the optimal strategy $q^\circ(\theta)$ of investors of type θ , and their indirect utility $v^\circ(\theta)$ are given by

$$(\phi^\circ(\theta), q^\circ(\theta), v^\circ(\theta)) = \left(\frac{2\zeta(\theta)}{\|\zeta(\theta)\|^2}, \frac{\|\zeta(\theta)\|^2}{4a}, \frac{\|\zeta(\theta)\|^2}{8} \right).$$

Our first result compares the risk exposures offered by the manager to a given type of investor the two models.

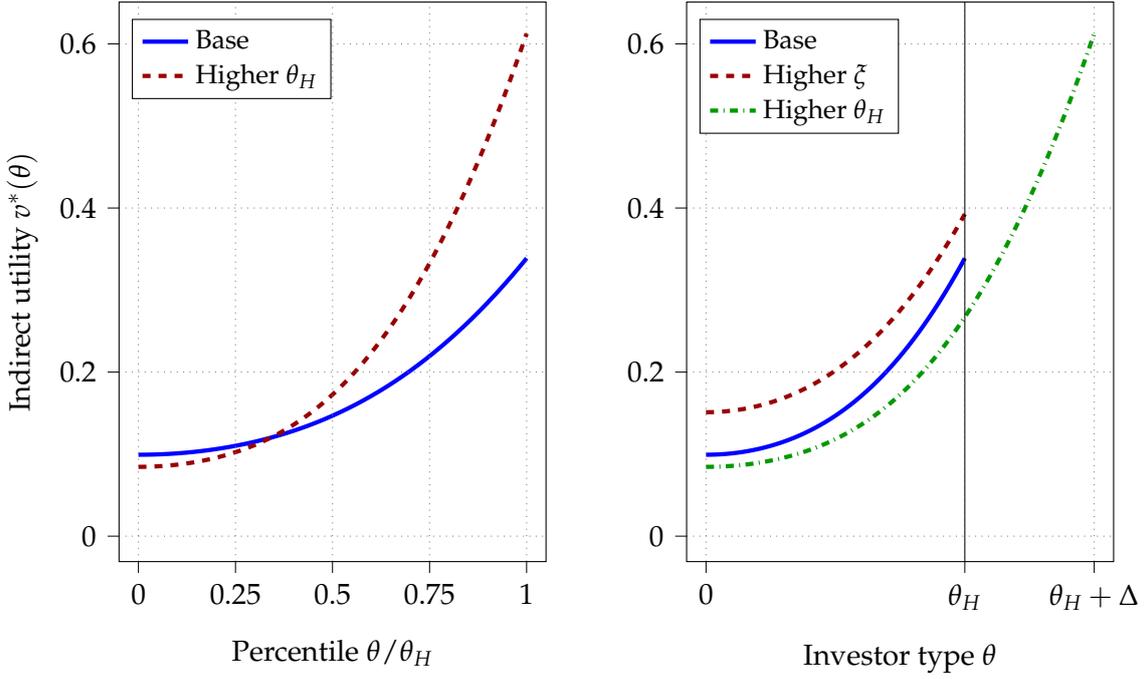


Figure 1: Effect of ζ and θ_H on the indirect utility of investors. The solid curve plots the investor's indirect utility as a function of his type (right) and his percentile in the distribution of types (left). The dashed curves illustrate how this utility changes when the range of perceived risk premia increases from θ_H to $\theta_H + \Delta$ while the dash-dotted curve indicates how it changes when the risk premium of the familiar asset increases.

Proposition 6 *The function*

$$\Delta(\theta) := \frac{\phi_2^\circ(\theta)}{\phi_1^\circ(\theta)} - \frac{\phi_2^*(\theta)}{\phi_1^*(\theta)} = \frac{\theta}{\zeta} - \frac{\check{v}^*(\theta)}{g^*(\theta)} \quad (14)$$

is nonnegative for all $\theta \in \Theta$ and such that $\Delta(0) = \Delta(\theta_H) = 0$.

The proposition shows that the lack of information regarding the beliefs of investors for the non-familiar asset leads the manager to offer funds that are more tilted towards the familiar asset than in the first best. This means that to achieve a given exposure in the non-familiar asset, a given investor needs to take a larger position in the familiar asset than he would have in the first best, and the manager uses the eagerness of investors to do so as a screening device. Our model thus provides a potential explanation for the fact that many mutual funds managers are perceived as being either closet indexers or home-biased, because their funds are too geared towards the familiar asset. Intuitively, the strength of this bias effect should be driven by the intensity of the information friction. Therefore, one expects that $\Delta(\theta)$ should increase for all types as θ_H/ζ increases, and Figure 2 confirms that this is indeed the case.

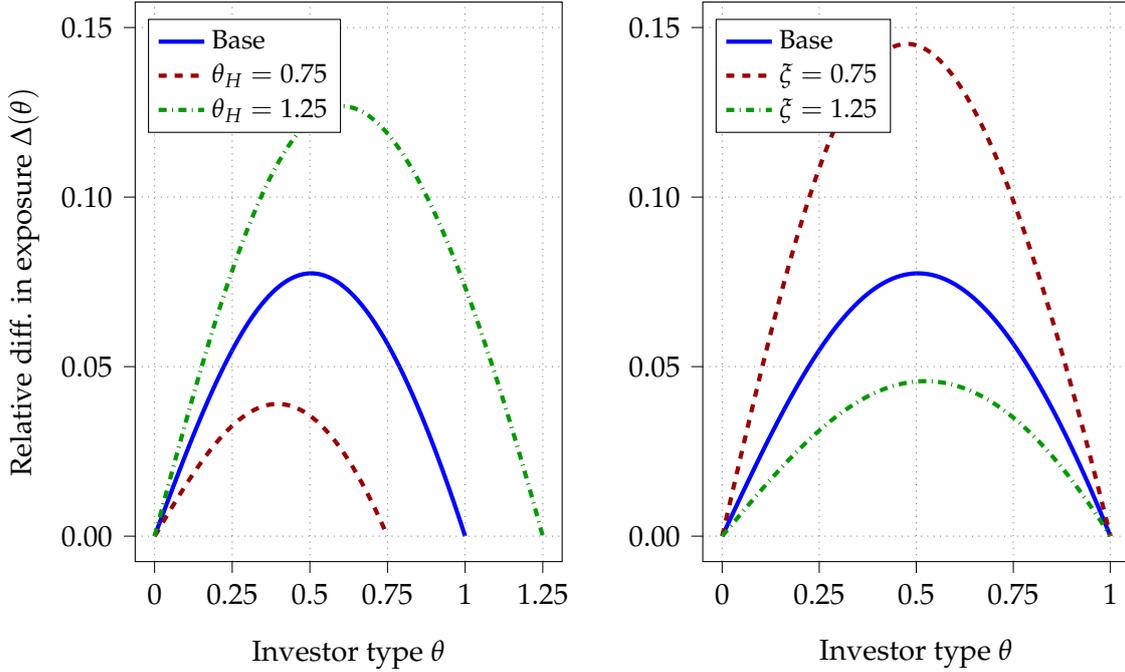


Figure 2: Relative difference in risk exposures compared to the first best. The solid curves plots the function $\Delta(\theta)$ defined in (14) at the base parameter values $\zeta = \theta_H = 1$. The dashed curves illustrate the impact of a decrease in either θ_H (left) or ζ (right) while the dash-dotted curves illustrate the impact of an increase in these same parameters.

The inverse U -shape of $\Delta(\theta)$ that is apparent in both panels of Figure 2 is the result of two conflicting effects. On the one hand, the fact that more optimistic investors demand more of the non-familiar asset prompts the manager to intensify the distortions as θ increases to deter these more optimistic investors from underreporting their type. On the other hand, as θ increases, the mass of investors who might be tempted to underreport their type as being equal to θ becomes smaller and this implies that fewer distortions are needed to maintain incentive compatibility.

The second part of Proposition 6 shows that the risk exposures offered to the most pessimistic and most optimistic investors are the same as in the first best. However, this does not imply that these investors select the same allocation or receive the same utility as in the first best because, even though the risk exposures they are offered are the same, the prices that the manager demands for them are different.

Our next result provides a detailed comparison of the portfolio allocations and indirect utilities of investors in the two models.

Proposition 7 *We have that*

$$\{\theta \in \Theta : v^*(\theta) \leq v^\circ(\theta)\} = [0, \bar{\theta}],$$

and

$$\{\theta \in \Theta : \pi(\theta, \phi^*(\theta))\phi_k^*(\theta) \leq q^\circ(\theta)\phi_k^\circ(\theta)\} = [0, \theta_k],$$

for $k \in \{1, 2\}$ and some types $\theta_1 \leq \bar{\theta} \leq \theta_2$.

The proposition shows that types below θ_1 are invested less in both risky assets than in the first best case and suffer a utility loss, that types above θ_2 are invested more in both risky assets than in the first best case and receive a utility gain, and that intermediate types in $[\theta_1, \theta_2]$ are invested more in the familiar asset and less in the non-familiar asset than in the first best case. To understand these results, start by considering investors with very low types. Since such investors lie at the bottom of the distribution the manager needs to deter almost all other investors from pooling with them. To do so he must offer them high prices and it naturally follows that such investors are invested less in both assets and suffer a significant utility loss compared to the first best benchmark. More optimistic agents are offered better terms that lead them to invest more and increase their exposure to the non-familiar asset. However, the discussion following Proposition 6 shows that in doing so their exposure to the familiar asset will increase at a faster rate which explains the appearance of an intermediate region where investors are invested more in the familiar asset and less in the non-familiar asset. Finally, as we approach the right tail of the distribution the terms that are being offered to investors are so good that they invest more in both assets, and their indirect utility exceeds that of the first best.

When moving from the first best benchmark to our asymmetric information case the manager naturally suffers a decrease in utility since he now has less information about the preferences of his clients. By Proposition 5 we have that his indirect utility is decreasing in θ_H and, since his first best indirect utility

$$\theta_H \mapsto v_m^\circ := \frac{1}{\theta_H} \int_0^\theta q^\circ(\theta) d\theta = \frac{1}{4a} \left(\xi^2 + \frac{1}{3}\theta_H^2 \right)$$

is increasing, we conclude that the manager's utility loss increases as the range of investor beliefs broadens. The mechanism behind this result is clear: As θ_H increases, investors become more optimistic on average and thus more eager to trade the non familiar asset. Therefore, the information friction becomes more intense and it follows that the manager's utility loss increases. Similarly, as the risk premium of the familiar asset increases the information friction becomes less intense because investors now

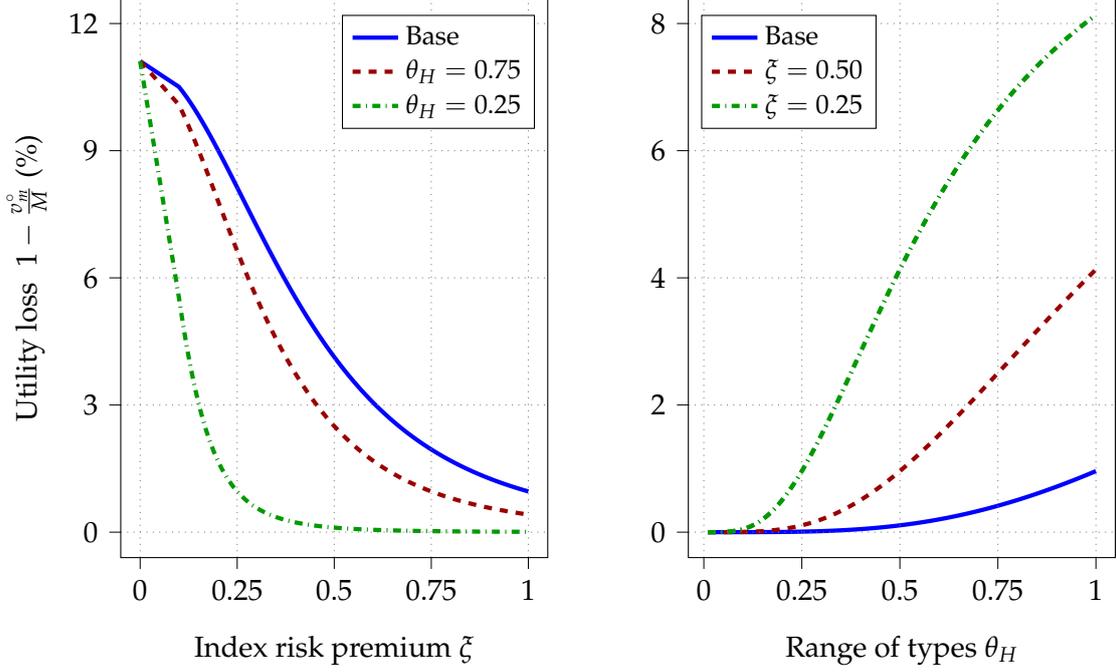


Figure 3: Manager's utility loss compared to the first best. The solid curves plot the relative utility loss of the manager as a function of the risk premium of the familiar asset and the range of perceived risk premia on the non-familiar asset. In each panel the dashed and dash-dotted curves illustrate the effect of a change in either the risk premium ζ of the familiar asset (right) or the range of investor types θ_H (left). The base case parameters used to construct this figure are $\zeta = \theta_H = 1$.

tend to care more for the familiar asset. As a result, we naturally expect the manager's utility loss to decrease as a function of ζ and the right panel of Figure 3 numerically confirms that this property holds at the optimum of our model.

2.3.1 Comparison to nonlinear pricing

To illustrate the impact of linear pricing we compare the outcomes of our model to those of the model in which the manager is unconstrained in his choice of the pricing scheme. We show in Appendix A that in this case the optimal strategy consists in offering a fixed cost of $\frac{1}{2a}\zeta^2$ for unconstrained access to the familiar asset and a quantity-dependent unit price given by

$$\hat{p}(q) := \frac{1}{2}\theta_H - \frac{a}{4}q. \quad (15)$$

for trading the non-familiar asset. In response to this menu, an investor of type $\theta \in \Theta$ demands $\hat{q}_1(\theta) = \frac{1}{a}\zeta$ units of the familiar asset and $\hat{q}_2(\theta) = \frac{1}{a}(2\theta - \theta_H)_+$ units of the

non-familiar asset so that his expected utility is given by

$$\hat{v}(\theta) := u \left(\theta, rw_0 + \hat{q}(\theta)^\top \epsilon - \frac{1}{2a} \bar{\zeta}^2 - \hat{q}_2(\theta) \hat{p}(\hat{q}_2(\theta)) \right) = \left(\theta - \frac{\theta_H}{2} \right)_+^2.$$

Comparing this solution to that of our model reveals two major differences. First, with nonlinear pricing investors of type $\theta = 0$ who only care about the familiar asset get zero utility which means that, in contrast to the linear pricing case, the manager is able to extract the whole surplus generated by investments in the familiar asset. This is intuitive. Indeed, because it is common knowledge that investors have identical preferences regarding the familiar asset, the manager knows exactly how many units each investor would want to acquire and thus can set his fixed price so as to extract the surplus generated by this investment.

Second, and more importantly, the use of nonlinear pricing makes it optimal for the manager to exclude investors who are less optimistic than average from the non-familiar asset market. By contrast, under linear pricing the optimal menu is such that all investors hold the two risky assets and receive a strictly positive utility. This suggests that linear pricing improves the aggregate welfare of investors and may even result in individual gains for all investors if the benefit from recovering part of the surplus associated with the familiar asset is sufficient to offset the forgone quantity discounts implied by (15) on the non-familiar asset. These intuitive properties seem difficult to establish analytically. However, all our numerical simulations confirm that linear pricing indeed improves the aggregate welfare of investors, and Figure 4 illustrates that it may even lead to strict Pareto improvements when the ratio θ_H/ζ that measures the intensity of the information friction is low enough.

2.3.2 The optimality of bundling

The normalization that we adopted in Proposition 1 implies that funds differ only in their exposure to the risky assets. This normalization is convenient for the derivation of the optimal menu but, in some cases, it may be more natural to instead normalize the funds in such a way that the optimal menu only includes all-equity funds that do not invest in the riskless asset. With this normalization the fund that targets investors

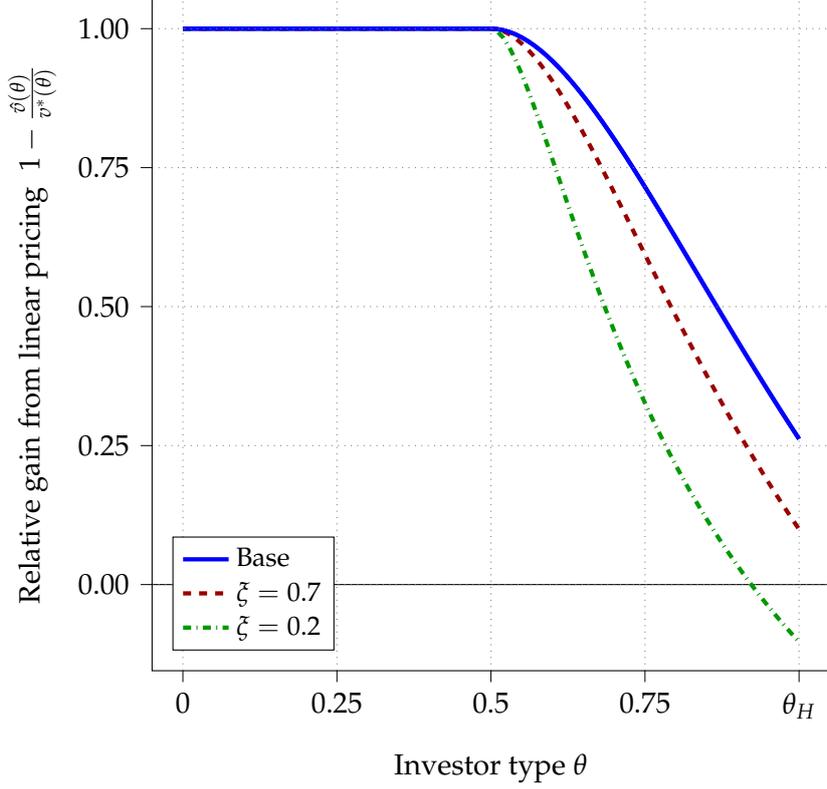


Figure 4: Utility gain from linear pricing. This figure plots the relative difference between the indirect utility of investors in the benchmark model and their indirect utility in the model where the manager can use nonlinear pricing. The solid curve represents gain from linear pricing in the base case where $\zeta = \theta_H = 1$ while the dashed and dash-dotted curves illustrate the impact of a gradual decrease in the familiar asset risk premium.

of type $\theta \in \Theta$ is given by

$$(\gamma_{AE}(\theta), \phi_{AE}(\theta)) := \frac{[1, \phi^*(\theta)]}{\phi_1^*(\theta) + \phi_2^*(\theta)} = \frac{1}{\dot{v}^*(\theta) + g^*(\theta)} \left[aF^*(\theta), \begin{pmatrix} g^*(\theta) \\ \dot{v}^*(\theta) \end{pmatrix} \right],$$

where we have set

$$g^*(\theta) := \sqrt{2v^*(\theta) - [\dot{v}^*(\theta)]^2},$$

In particular, since $\dot{v}^*(0) = 0$ by Theorem 1, we have $\phi_{AE}(0) = (1, 0)^\top$ so that the familiar asset is offered in the optimal menu with a fee rate given by

$$\gamma_1^* := \frac{aF^*(0)}{g^*(0)} = \zeta - \sqrt{2v^*(0)} \in \left[\frac{1}{3}, \frac{2}{3} \right] \zeta, \quad (16)$$

where the inclusion follows from the fact that, as we show in Appendix 2.3, the indirect utility of the most pessimistic investor lies in the interval $[\frac{1}{18}, \frac{2}{9}]\zeta^2$.

This shows that the requirement of linear pricing leads the manager to engage in what Adams and Yellen (1976) refer to as *mixed* bundling — the familiar asset is available both separately and in packages. Note, however, that it is never optimal for the manager to also offer the non-familiar asset separately because under a linear pricing constraint the most effective way for him to screen investors is to bundle the assets. Indeed, if the manager decides not to bundle, then the best he can do is to offer the two assets separately with fee rates given by

$$(\hat{\gamma}_1, \hat{\gamma}_2) := \arg \max_{\gamma \in \mathbf{R}_+^2} \left\{ \gamma_1 (\zeta - \gamma_1)_+ + \frac{1}{\theta_H} \int_0^{\theta_H} \gamma_2 (\theta - \gamma_2)_+ d\theta \right\} = \left(\frac{\zeta}{2}, \frac{\theta_H}{3} \right). \quad (17)$$

This offering in turn generates

$$M_0 := \hat{\gamma}_1 (\zeta - \hat{\gamma}_1)_+ + \frac{1}{\theta_H} \int_0^{\theta_H} \hat{\gamma}_2 (\theta - \hat{\gamma}_2)_+ d\theta = \frac{1}{4}\zeta^2 + \frac{2}{27}\theta_H^2$$

in management fees, and our next result confirms that this quantity is strictly lower than the amount of fees generated by the optimal fund menu.

Lemma 2 *It is never optimal for the manager to unbundle the assets, that is $M_0 < M$.*

2.3.3 Exclusivity of funds

As discussed in Section 2.1.2, the first order conditions induced by the exclusive and non-exclusive incentive compatibility constraints coincide. Therefore, the manager's indirect utility satisfies $M \leq M_E \leq V/\theta_H$ where

$$M_E := \sup \left\{ I(\phi) : \phi \in AC(\Theta; \mathbf{R}^2) \text{ such that (6) holds} \right\}$$

denotes his indirect utility when he can enforce exclusivity and it thus follows from Theorem 1 that $M = M_E$. Even if she had the ex-ante ability to commit investors to a single fund, the manager would not optimally offer funds that investors would want to combine if they could. As we will see below in Section 3.3 this result no longer holds when investors can directly access the familiar asset at a sufficiently low cost.

3 Direct investment in the familiar asset

In the benchmark model of Section 1 investors can only access the risky assets through the mutual funds offered by the manager. We now relax this assumption by allowing them to directly access the first risky asset by investing in an outside fund that only loads on the familiar asset risk.

3.1 Incentive compatible menus

Assume that investors can allocate their capital to the riskless asset, the funds offered by the manager, and a fund with gross excess return $R_1 := \epsilon_1 - \gamma_1$ where the constant $\gamma_1 \in [0, \xi)$ is the fee rate for direct investment in the familiar asset. In this case the action set of investors is $\mu_+(\mathcal{M}) \times \mathbf{R}_+$ and we associate to each element (q, n) of this set the terminal wealth

$$w_1(q, n, \mathbf{m}) := rw_0 + \int_{\mathcal{M}} R(\gamma(m), \phi(m)) q(dm) + nR_1.$$

In accordance with the model set forth in Section 1 the optimization problem of an investor of type $\theta \in \Theta$ is then defined by

$$v_i(\theta, \mathbf{m}) := \sup_{(q, n) \in \mu_+(\mathcal{M}) \times \mathbf{R}_+} u(\theta, w_1(q, n, \mathbf{m})),$$

and the aggregation of the investors' decisions generates the amount of management fees given by (2) where

$$(q^*(m; \theta, \mathbf{m}), n^*(\theta, \mathbf{m})) = \arg \max_{(q, n) \in \mu_+(\mathcal{M}) \times \mathbf{R}_+} u(\theta, w_1(q, n, \mathbf{m}))$$

denotes the best response of an investor of type $\theta \in \Theta$ to the menu offered by the manager. To allow for a simple analysis of the problem, we assume throughout this section that when the manager decides to include the familiar asset in his menu the with a fee that is equal to that which investors have to pay to trade on their own, then investors will direct their demand for the familiar asset to the manager rather than to the outside fund. Given this assumption, a fund menu \mathbf{m}^* is said to be optimal if it maximizes the total amount of management fees in (2).

The following result is the analogue of Proposition 1 for the case in which investors have direct access to the familiar asset. The only difference is Property 3 which is

an incentive compatibility condition that takes into account the possibility of direct investment in the familiar asset by requiring that an investor of type $\theta \in \Theta$ finds it optimal to invest the amount $\pi(\theta, \phi(\theta))$ in the fund $(1, \phi(\theta))$ that specifically targets his type, and nothing in either the outside fund or any of the other funds in the menu.

Proposition 8 *Assume that investors can directly access the familiar asset. Then, given any fund menu $\bar{\mathbf{m}}$ there exists a fund menu $\mathbf{m} = (\gamma, \phi, \mathcal{M})$ such that*

1. $\mathcal{M} = \Theta$,
2. $\gamma(\theta) = 1$ for all $\theta \in \Theta$,
3. $q^*(\sigma; \theta, \mathbf{m}) = \mathbf{1}_{\{\theta \in \sigma\}} \pi(\theta, \phi(\theta))$ and $n^*(\theta; \mathbf{m}) = 0$ for all $\theta \in \Theta$ and $\sigma \subseteq \Theta$,
4. $v_m(\mathbf{m}) = v_m(\bar{\mathbf{m}})$, and
5. $v_i(\theta, \mathbf{m}) = v_i(x, \bar{\mathbf{m}})$ for all $\theta \in \Theta$.

In view of the above result we have that the manager's problem can be reduced to the maximization of the integral $I(\phi)$ in (3) over the set $\Phi_{0,I}$ of fund loading functions that satisfy the incentive compatibility condition of Proposition 8. Our next result provides an explicit characterization of this set.

Proposition 9 *A fund loading function $\phi : \Theta \rightarrow \mathbf{R}^2$ is incentive compatible given direct access to the familiar asset if and only if it satisfies (4) and*

$$\zeta - \gamma_1 \leq \frac{\phi_1(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \zeta(\theta) - 1 \right)_+ \quad (18)$$

for all types $\theta \in \Theta$.

The interpretation of the above result is similar to that of Proposition 2, but there is an additional restriction: Since investors now have access to the familiar asset the manager also needs to make sure that the familiar asset has a negative alpha relative to any of the funds he offers, for otherwise some investors may find it optimal to combine the outside fund with the funds in the menu. Because $\zeta - \gamma_1 > 0$ this additional restriction implies that a necessary condition for incentive compatibility is that each fund in the menu provides a strictly positive expected gross excess return to the investors that it targets. Indeed, if this condition fails, then the right hand side of (18) will be zero for some $\theta \in \Theta$ and, as a result, investors of these types will find it optimal to invest exclusively in the outside fund rather than in any of the funds offered by the manager.

3.2 The relaxed problem

Propositions 8 and 9 imply that the manager's problem now consists in maximizing (3) over the set $\Phi_{1,0}$ of fund loading functions that satisfy (4) and (18). To solve that problem we further restrict the manager's choice by requiring that the fund loading function that he selects belongs to the set $\Phi_1 := \Phi_{1,0} \cap AC(\Theta; \mathbf{R}^2)$. The optimization problem that we consider is therefore given by

$$M_1 := \sup_{\phi \in \Phi_1} I(\phi). \quad (\mathcal{P}_1)$$

Before proceeding with the formulation of the appropriate relaxed problem, we start by identifying conditions under which having a direct access to the familiar asset is worthless to investors in the sense that the solution to the above problem coincides with the optimal fund loading function that we derived in the benchmark case.

Lemma 3 *Assume that the fee rate on the familiar asset is such that*

$$\gamma_1 > \gamma_1^* := \bar{\zeta} - \sqrt{2v^*(0)}$$

where the function v^ is defined as in Proposition 3. Then $M_1 = M$ and the optimal fund menu is given by Theorem 1.*

The intuition for the above lemma is simple. When confronted with two funds that offer the same risk exposure, but different fee rates, investors will systematically discard the fund with the higher fee rate. Therefore, if the fee rate γ_1 on the outside fund exceeds the fee rate γ_1^* that the manager offers as part of the optimal menu in the benchmark model of Section 1, then investors will optimally stay away from the outside fund to which they have direct access and the optimal fund menu will remain the same as in the benchmark model.

Assume from now on that the condition of Lemma 3 fails, so that having a direct access to the familiar asset is potentially valuable to investors. Following the logic of Lemma 1, our next result shows that the manager can use the indirect utility and marginal utility of investors as instruments.

Lemma 4 *Assume that $\phi \in \Phi_1$ is incentive compatible. Then, the indirect utility function defined by (5) and (6) belongs to the space $AC(\Theta; \mathbf{R})$, and satisfies both*

$$2v(\theta) \geq (\bar{\zeta} - \gamma_1)^2 + [\dot{v}(\theta)]^2 \quad (19)$$

and (8) for almost every type $\theta \in \Theta$.

Relying on the above lemma we have that the value of the manager's problem optimization problem satisfies

$$\theta_H M_1 \leq V_1 := \sup_{v \in \mathcal{V}_1} \int_{\Theta} F(\theta, v(\theta), \dot{v}(\theta)) d\theta \quad (\mathcal{R}_1)$$

where \mathcal{V}_1 denotes the set of functions $v \in AC(\Theta; \mathbf{R})$ that satisfy (19). An important difference between this relaxed problem and the one we formulated in the benchmark model is that for (19) to hold it is no longer sufficient that the integrand in (\mathcal{R}_1) be real valued for all types. As a result, the usual solution technique based on the Euler-Lagrange equation cannot be applied, and we will need to derive a set of optimality conditions that explicitly take into account the constraint. To do so, consider the Lagrangian objective

$$\int_{\Theta} H^\lambda(\theta, v(\theta), \dot{v}(\theta)) d\theta := \int_{\Theta} F(\theta, v(\theta), \dot{v}(\theta)) d\theta + \int_{\Theta} \lambda(\theta) c(v(\theta), \dot{v}(\theta)) d\theta$$

where the function

$$c(v(\theta), \dot{v}(\theta)) := 2v(\theta) - [\dot{v}(\theta)]^2 - (\xi - \gamma_1)^2$$

returns the value of the constraint (19) and $\lambda : \Theta \rightarrow \mathbf{R}_+$ is a Lagrange multiplier that enforces this constraint at each point of the type space.

The next lemma provides a set of sufficient optimality conditions for the relaxed problem (\mathcal{R}_1) . To state the result, denote by $AC_p^*(\Theta; \mathbf{R})$ the set of piecewise absolutely continuous functions that are right continuous at zero, left continuous at θ_H , and have at most finitely many jump discontinuities.

Lemma 5 *Let $(v, \lambda) \in \mathcal{V}_1 \times AC_p^*(\Theta; \mathbf{R}_+)$ be such that $\dot{v} \in AC(\Theta; \mathbf{R})$ and denote by \mathcal{C} the set of points where the function λ is continuous. If*

$$\left(H_{v(\theta)}^\lambda - \frac{d}{d\theta} H_{\dot{v}(\theta)}^\lambda \right) (\theta, v(\theta), \dot{v}(\theta)) = 0, \quad \theta \in \mathcal{C}, \quad (20)$$

$$H_{\dot{v}(\theta)}^\lambda (\theta, v(\theta), \dot{v}(\theta)) = 0, \quad \theta \in \{0, \theta_H\}, \quad (21)$$

$$\lambda(\theta) c(v(\theta), \dot{v}(\theta)) = 0, \quad \theta \in \Theta, \quad (22)$$

and $H_{\dot{v}(\theta)}^\lambda (\theta, v(\theta), \dot{v}(\theta))$ is continuous, then v attains the supremum in (\mathcal{R}_1) .

The above conditions can be interpreted as follows. The first condition requires that the Euler-Lagrange equation associated with the optimization of the Lagrangian holds at all points of continuity of the multiplier. The second condition imposes the boundary conditions associated with the free values of the curve on the boundaries of the type space. The third condition is the usual complementary slackness condition associated with the optimal choice of the multiplier.

The last condition is technical. It is meant to provide a sufficient condition for the integration by parts argument that proves the optimality of the candidate derived from the first three conditions, and is thus similar to the first corner condition of Weierstrass (see, e.g., [Mesterton-Gibbons \(2009, Chapter 6\)](#)). In the context of our problem this condition requires that the function

$$H_{\dot{v}(\theta)}^\lambda(\theta, v(\theta), \dot{v}(\theta)) = \theta - \dot{v}(\theta) \left(2\lambda(\theta) + \frac{\dot{v}(\theta)\xi}{\sqrt{2v(\theta) - [\dot{v}(\theta)]^2}} \right)$$

be continuous throughout the type space, and, since the functions v and \dot{v} are both continuous by assumption, this requirement is equivalent to the property that $\dot{v}(\theta) = 0$ at every point of discontinuity of the Lagrange multiplier.

To construct a pair (v, λ) satisfying the optimality conditions of Lemma 5 we need to conjecture the shape of the region in which the constraint binds. Investors with low types do not care much about the non-familiar asset and have access to the familiar asset at a relatively cheap price. To entice such investors to allocate capital to a fund that includes the non-familiar asset, the manager needs to offer them very favorable conditions, but the resulting fund is likely to fail the incentive compatibility constraint. As a result, we expect the manager to offer the familiar to all investors below a certain cutoff type and it follows that the constraint should bind over $[0, \theta^*]$ for some threshold $\theta^* \in \Theta$. As the fee rate on the familiar asset decreases all investors become less willing to acquire exposure to the familiar asset otherwise than through the outside fund. This makes it more difficult for the manager to screen investors by bundling and we therefore expect that below a certain fee rate the manager will no longer find it optimal to do so. At that point the manager will pick a menu that is equivalent to offering the two assets separately with the fee rates given by γ_1 for the familiar asset $\hat{\gamma}_2 = \frac{1}{3}\theta_H$ for the non-familiar asset (see (17)). This menu delivers the indirect utility

$$s(\theta) := \sup_{q \in \mathbf{R}_+^2} u\left(\theta, rw_0 + q^\top(\epsilon - \gamma)\right) = \frac{1}{2}(\xi - \gamma_1)^2 + \frac{1}{2}\left(\theta - \frac{\theta_H}{3}\right)_+^2, \quad (23)$$

and a direct calculation shows that this indirect utility saturates the constraint (19). Based on this observation we conjecture that when γ_1 is sufficiently low, the constraint binds not only for low types but throughout the type space.

The following proposition confirms the above conjectures and provides a complete solution to the relaxed problem.

Proposition 10

- a) If $\gamma_1 \leq \frac{1}{3}\xi$, then the function $v_1^*(\theta) := s(\theta)$ attains the supremum in (\mathcal{R}_1) .
- b) Assume $\gamma_1 \in (\frac{1}{3}\xi, \gamma_1^*]$ and denote by $(w, \theta^*) \in C_p^2(\Theta; \mathbf{R}) \times \Theta$ the unique solution to the free boundary problem defined by

$$w(\theta) (1 + \ddot{w}(\theta)) = [\dot{w}(\theta)]^2 + \frac{3}{2\xi} \left(2w(\theta) - [\dot{w}(\theta)]^2 \right)^{\frac{3}{2}} \quad (24)$$

subject to

$$0 = \dot{w}(\theta^*) = w(\theta^*) - \frac{1}{2} (\xi - \gamma_1)^2 \quad (25)$$

$$= \theta_H - \xi \dot{w}(\theta_H) \left(2w(\theta_H) - [\dot{w}(\theta_H)]^2 \right)^{-\frac{1}{2}}. \quad (26)$$

Then, the function defined by

$$v_1^*(\theta) := \frac{1}{2} (\xi - \gamma_1)^2 + \mathbf{1}_{\{\theta > \theta^*\}} \left(w(\theta) - \frac{1}{2} (\xi - \gamma_1)^2 \right) \quad (27)$$

attains the supremum in (\mathcal{R}_1) .

3.3 The optimal fund menu

As in the benchmark case, we have that the value of the relaxed problem (\mathcal{R}_1) gives an upper bound on the value of the manager's problem (\mathcal{P}_1) . To show that these two values coincide, we need to construct a fund loading function $\phi_1^* \in \Phi_1$ that delivers to each investor the indirect utility prescribed for his type by the solution to the relaxed problem. This is the content of the following:

Theorem 2 Assume $\gamma_1 \leq \gamma_1^*$ and denote by $v_1^* \in C_p^2(\Theta; \mathbf{R})$ the solution to the relaxed problem (\mathcal{R}_1) as defined in Proposition 10. Then

$$\inf_{\theta \in \Theta} F(\theta, v_1^*(\theta), \dot{v}_1^*(\theta)) > 0, \quad (28)$$

and the fund loading function

$$\phi_1^*(\theta) := \frac{1}{aF(\theta, v_1^*(\theta), \dot{v}_1^*(\theta))} \left(\sqrt{2v_1^*(\theta) - [\dot{v}_1^*(\theta)]^2}, \dot{v}_1^*(\theta) \right)^\top \quad (29)$$

attains the supremum in (\mathcal{P}_1) . In particular, $\gamma_1 \phi_1^*(\theta) = (1, 0)^\top$ for all $\theta \leq \theta^*$, so that the manager only offers the familiar asset to all sufficiently low types.

A comparison of Theorems 1 and 2 reveals two important differences. First, the presence of competition from the outside fund forces the manager to offer the familiar asset at the market rate γ_1 rather than at the monopolistic fee rate γ_1^* of Lemma 3. In addition, the optimal menu is such that the manager offers the familiar asset not only to the most pessimistic investors, but to a group of sufficiently pessimistic investors who, therefore, find themselves excluded from the non-familiar asset market.

Second, if the competition induced by the outside fund is sufficiently fierce, then it is no longer useful for the manager to bundle assets and the optimum can be implemented by simply offering the familiar asset at the market rate γ_1 and the non-familiar asset at rate $\frac{1}{3}\theta_H$. Note however that this conclusion is fragile and dependent on our specific assumptions. In our model the introduction of a cheap outside fund effectively reduces the optimal number of funds from a continuum to only two but this only occurs because there are only two risky assets. If there were more, then it would most likely remain optimal for the manager to offer of a continuum of funds irrespective of the fee rate on the familiar asset.

Since the market fee rate γ_1 on the familiar asset is lower than the monopolistic fee rate γ_1^* , it follows from (16) that we have

$$v_1^*(0) = \frac{1}{2}(\xi - \gamma_1)^2 \geq \frac{1}{2}(\xi - \gamma_1^*)^2 = v^*(0).$$

This shows that the presence of an outside fund improves the welfare of the most pessimistic investors who care the most for the familiar asset. As illustrated by Figure 5, this property actually holds for all investors because the presence of a cheap outside fund combined with the need to maintain incentive compatibility forces the manager to offer better terms not only to the most pessimistic investors but to all of them.

To conclude let us briefly examine the non exclusivity of funds in the presence of an outside fund. If the manager has the ability to commit each investor to a single fund, then incentive compatibility requires that the indirect utility of investors satisfies (6)

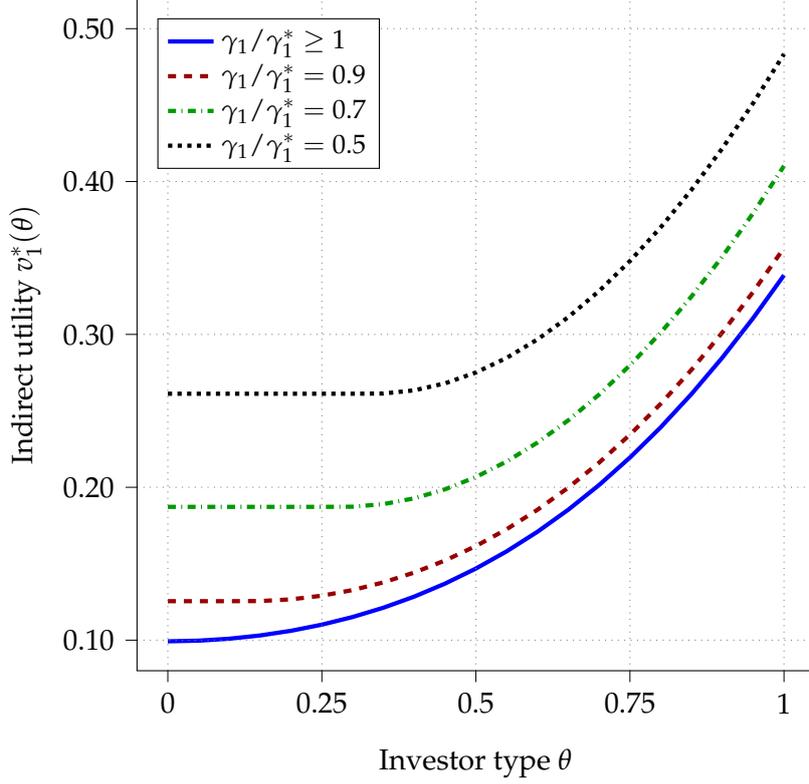


Figure 5: Indirect utility in the presence of an outside fund. This figure plots the indirect utility of an investor as a function of his type for different levels of the market fee rate γ_1 on the outside fund.

and a participation constraint that is now given by

$$\inf_{\theta \in \Theta} v(\theta) \geq \frac{1}{2}(\bar{\xi} - \gamma_1)^2, \quad (30)$$

which simply states that optimally picking one fund out of the menu is preferable to optimally investing in the outside fund. Our last result provides a complete solution to the model in this case and shows that, contrary to what happens in the benchmark case, the ability to commit investors to a single fund can have value when some measure of competition is included in the model.

Proposition 11 *Assume that the manager can commit each investor to a single fund. Then, the indirect utility of investors and the optimal fund loading function are explicitly given by (27) and (29) for all $\gamma_1 \leq \gamma_1^*$. In particular, the ability to commit investors has strictly positive value to the manager if and only if $\gamma_1 < \frac{1}{3}\bar{\xi}$.*

4 Conclusion

In this paper we argue that offering a menu of funds is optimal for an investment firm that has incomplete information about the characteristics of its customer base and is required to use linear pricing. To illustrate this mechanism we study the optimal offering strategy of a manager who is constrained to use fraction-of-fund fees and does not observe the beliefs of investors regarding one of the risky assets. We show that the optimal menu can be explicitly constructed from the solution to a calculus of variations problem that optimizes over the indirect utility that investors receive. We provide a complete characterization of the optimal menu and study its most salient features. In particular, we show that the information friction leads the manager to offer funds that are inefficiently tilted towards the asset that is not subject to the information friction, and argue that this result provides a novel explanation for either closet indexing or the home bias depending on the interpretation given to the familiar asset.

While the tractability of our model rests on specific assumptions regarding the beliefs, preferences and heterogeneity of investors, many of our results are not dependent on these assumptions. Instead of (or in addition to) differing in their beliefs, the investors could also differ along other dimensions such as risk aversion, initial endowments, the assets they are willing to invest in, or the risks they are exposed to prior to choosing their fund allocation. Such generalizations are likely to complicate the solution to the model but they would not change our qualitative message. In particular, linear pricing would still induce the manager to bundle assets as a way of screening investors, and the need to maintain incentive compatibility would still generate a loss of efficiency compared to the full information benchmark. We leave these challenging extensions of our basic framework for future research.

Optimal fund menus: Appendix

Jakša Cvitanić Julien Hugonnier

October 29, 2018

This document presents the solution to various benchmark versions of our model and gathers the proofs of all the results in the main text.

A Benchmarks	1
A.1 Free access to all assets	1
A.2 Nonlinear pricing	1
A.3 Linear pricing under complete information	3
B Proof of the results in Section 2	5
B.1 The revelation principle	5
B.2 The relaxed problem	10
B.3 The optimal fund menu	21
C Proof of the results in Section 3	28
C.1 Incentive compatible menus	28
C.2 The relaxed problem	28
C.3 The optimal fund menu	33

*First version: July 2017. This work was initiated during a visit of Jakša Cvitanić to EPFL in the context of the special semester on *Stochastic Dynamic Models in Mathematical Finance, Econometrics, and Actuarial Sciences* organized by the [Centre International Bernoulli](#) in the Spring of 2017. We thank Jonathan Berk, Ryan Donnelly, Federico Echenique, Ruediger Fahlenbrach, Damir Filipovic, Benjamin Hebert, Daniel Kuhn, Semyon Malamud, Yaroslav Melnyk, Erwan Morellec, Chris Rogers, and seminar participants at Cambridge, USC, and CalTech for conversations on the topic of this paper.

†Division of the Humanities and Social Sciences, CalTech. M/C 228–77, 1200 East California Blvd., Pasadena CA 91125, USA. Email: cvitanic@hss.caltech.edu.

‡EPFL, Swiss Finance Institute, and CEPR. Extranef 212, Quartier UNIL Dorigny, 1015 Lausanne, Switzerland. Email: julien.hugonnier@epfl.ch.

A Benchmarks

In this appendix we briefly present the solution of our model in three benchmark cases: the case in which investors can freely trade all the existing assets, the case in which the manager does not observe the types of investors, but he is allowed to offer any price schedules, and finally the case in which the manager has full information about the types of investors, but he is required to use linear price schedules.

A.1 Free access to all assets

Assume that investors can freely trade all existing assets. In this case the indirect utility of the manager is zero, and an investor of type $\theta \in \Theta$ chooses his optimal investment in the two risky asset by solving

$$\tilde{v}(\theta) := \sup_{q \in \mathbf{R}^2} u(\theta, rw_0 + R(0, q)) = \sup_{q \in \mathbf{R}^2} \left\{ aq^\top \xi(\theta) - \frac{a^2}{2} \|q\|^2 \right\}.$$

The optimal solution to this concave problem is $\tilde{q}(\theta) := \frac{1}{a} \xi(\theta)$. Substituting this solution back into the objective function shows that the investor's indirect utility is

$$\tilde{v}(\theta) = a\tilde{q}(\theta)^\top \xi(\theta) - \frac{a^2}{2} \|\tilde{q}(\theta)\|^2 = \frac{1}{2} \|\xi(\theta)\|^2.$$

This function naturally constitutes an upper bound on the indirect utility that investors can hope to achieve when they no longer have direct access to all assets.

A.2 Nonlinear pricing

Assume now that the manager does not observe the types of investors, but is allowed to use any price schedule that is a function of the amount invested. In this case an investor of type $\theta \in \Theta$ chooses his allocation by solving

$$\hat{v}(\theta) := \sup_{q \in \mathbf{R}_+^2} u(\theta, rw_0 + R(0, q) - P(q))$$

where the function $P(q)$ is the fee that the manager charges for a portfolio $q \in \mathbf{R}_+^2$. Denoting the solution to this problem by $\hat{q}(\theta)$, we have that the manager's problem

can be formulated as

$$\hat{M} := \sup_p \int_{\Theta} p(\hat{q}(\theta)) \frac{d\theta}{\theta_H}.$$

Since

$$p(\hat{q}(\theta)) = \hat{q}(\theta)^\top \xi(\theta) - \frac{a}{2} \|\hat{q}(\theta)\|^2 - \frac{\hat{v}(\theta)}{a} \quad (31)$$

we have that the manager's objective function can be written as

$$\begin{aligned} \int_{\Theta} p(\hat{q}(\theta)) \frac{d\theta}{\theta_H} &= \int_{\Theta} \left(\hat{q}(\theta)^\top \xi(\theta) - \frac{a}{2} \|\hat{q}(\theta)\|^2 - \frac{\hat{v}(\theta)}{a} \right) \frac{d\theta}{\theta_H} \\ &= \int_{\Theta} \left(\hat{q}(\theta)^\top \xi(\theta) - \frac{a}{2} \|\hat{q}(\theta)\|^2 - \frac{1}{a} \frac{d\hat{v}(\theta)}{d\theta} (\theta_H - \theta) \right) \frac{d\theta}{\theta_H} - \frac{\hat{v}(0)}{a} \\ &= \int_{\Theta} \left(\hat{q}_1(\theta) \xi + \hat{q}_2(\theta) (2\theta - \theta_H) - \frac{a}{2} \|\hat{q}(\theta)\|^2 - \frac{\hat{v}(\theta)}{a} \right) \frac{d\theta}{\theta_H}, \end{aligned}$$

where the second equality follows from integration by parts, and the third follows from the envelope condition which requires that

$$\frac{d\hat{v}}{d\theta}(\theta) = \frac{d}{d\theta} \left\{ a \left(q^\top \xi(\theta) - P(q) \right) - \frac{a^2}{2} \|q\|^2 \right\}_{q=\hat{q}(\theta)} = a\hat{q}_2(\theta).$$

Maximizing under the integral sign shows that whenever possible the manager should choose the price function in such a way that $\hat{v}(0) = 0$ and

$$\hat{q}(\theta) = \frac{\mathbf{e}_1}{a} \xi + \frac{\mathbf{e}_2}{a} (2\theta - \theta_H)_+$$

where $(\mathbf{e}_1, \mathbf{e}_2)$ denotes the orthonormal basis of \mathbf{R}^2 . Since the above allocation is weakly increasing, it follows from well-known results in the screening literature (see for example [Laffont and Martimort \(2009\)](#)) that there exists a price function $\hat{P}(q)$ that implements it, in the sense that

$$\hat{q}(\theta) \in \arg \max_{q \in \mathbf{R}_+^2} \left\{ a \left(q^\top \xi(\theta) - \hat{P}(q) \right) - \frac{a^2}{2} \|q\|^2 \right\}.$$

The indirect utility of an investor of type $\theta \in \Theta$ can be computed from the envelope condition which requires that

$$\hat{v}(\theta) = \hat{v}(0) + \int_0^\theta d\hat{v}(x) = \int_0^\theta a\hat{q}_2(x)dx = \left(\theta - \frac{1}{2}\theta_H\right)_+^2.$$

Using this formula in conjunction with (31) then shows that the amount of fees that the managers receives from an investor of type $\theta \in \Theta$ is

$$\hat{P}(\hat{q}(\theta)) = \frac{1}{2a}\zeta^2 + \frac{1}{a}(2\theta - \theta_H)_+ \left(\frac{3}{4}\theta_H - \frac{1}{2}\theta\right)$$

and it now follows from the taxation principle (see [Rochet \(1985\)](#)) that a price function which allows the manager to achieve his optimum is given by

$$\hat{P}(q) = \mathbf{1}_{\{q_1 \neq 0\}} \frac{1}{2a}\zeta^2 + q_2 \left(\frac{1}{2}\theta_H - \frac{a}{4}q_2\right).$$

The interpretation of the above results is clear. All investors are able to achieve their optimal level of exposure to the familiar asset (i.e. $\hat{q}_1(\theta) = \tilde{q}_1(\theta)$), but the fact that manager is fully informed about investors' preferences regarding the familiar asset allows him to fully extract the corresponding surplus. On the other hand, because the manager does not observe the investors' preferences regarding the non-familiar asset, he must screen them along this dimension. To this end, he uses a price schedule for exposure to the non-familiar asset that is strictly concave so that marginal prices decrease with quantities. In equilibrium, all investors except those of the highest type achieve inefficiently small exposures to the non-familiar asset, and investors who are less optimistic than average even get excluded from that market because serving such investors would require lower prices that would in turn reduce the fees that the manager extracts from more optimistic investors.

A.3 Linear pricing under complete information

Assume now that the manager is fully informed about the type of each investor and interacts with each of them in a bilateral way, but is still required to use linear price schedules. In this case the solution of the model can be constructed by solving for the single fund that the manager will offer to investors of a given type $\theta \in \Theta$. If the manager offers a fund (γ, ϕ) to such an investor, then the amount that this investor

will optimally allocate to the fund is given by

$$\begin{aligned} \arg \max_{q \in \mathbf{R}_+} u(\theta, rw_0 + qR(\gamma, \phi)) &= \arg \max_{q \in \mathbf{R}_+} \left\{ aq \left(\phi^\top \xi(\theta) - \gamma \right) - \frac{a^2}{2} \|q\phi\|^2 \right\} \\ &= \frac{1}{a\|\phi\|^2} \left(\phi^\top \xi(\theta) - \gamma \right)_+, \end{aligned} \quad (32)$$

where the second equality follows from an application of the Kuhn-Tucker conditions. Taking this best response into account the manager then solves

$$v_m^\circ(\theta) := \sup_{(\gamma, \phi) \in \mathbf{R} \times \mathbf{R}^2} \frac{\gamma}{a\|\phi\|^2} \left(\phi^\top \xi(\theta) - \gamma \right)_+.$$

Since the objective function of this problem only depends on the vector $v = \phi/\gamma$, we may without loss of generality normalize the fee rate to 1. With this normalization the manager's problem boils down to

$$v_m^\circ(\theta) = \sup_{v \in \mathbf{R}^2} \frac{1}{a\|v\|^2} \left(v^\top \xi(\theta) - 1 \right)_+, \quad (33)$$

and solving that problem shows that the linearly priced fund that the manager offers to investors of type $\theta \in \Theta$ is given by $(1, \phi^\circ(\theta))$ with

$$\phi^\circ(\theta) := \arg \max_{v \in \mathbf{R}^2} \frac{1}{a\|v\|^2} \left(v^\top \xi(\theta) - 1 \right)_+ = \frac{2\xi(\theta)}{\|\xi(\theta)\|^2}.$$

Substituting this expression into (32) and (33) shows that the amount that investors of type $\theta \in \Theta$ allocate to the fund, their indirect utility, and the manager's indirect utility are explicitly given by

$$(q^\circ(\theta), v^\circ(\theta)) := \left(\frac{\|\xi(\theta)\|^2}{4a}, \frac{\|\xi(\theta)\|^2}{8} \right)$$

and

$$v_m^\circ := \int_{\Theta} v_m^\circ(\theta) \frac{d\theta}{\theta_H} = \frac{1}{4a} \left(\xi^2 + \frac{1}{3}\theta_H^2 \right).$$

The most salient features of this solution can be summarized as follows. First, and contrary to what happens when all price schedules are allowed, the indirect utility of investors depends on the risk premium ξ of the familiar asset. This shows that the requirement of linear pricing prevents the manager from extracting the whole surplus

associated with the familiar asset. Second, each investor allocates a strictly positive amount to the fund that the manager proposes to him. Third, the optimal exposure of an investor of type $\theta \in \Theta$ to the two risky assets are given by

$$q_2^\circ(\theta)\phi_2^\circ(\theta) = \frac{\bar{\zeta}(\theta)}{2a}.$$

Because investors have quadratic preferences and agree on the familiar asset risk premium, their optimal exposures to the familiar asset are the same, but more optimistic investors naturally choose a larger exposure to the non-familiar asset. The relative composition of the investors' optimal portfolios

$$\frac{q^\circ(\theta)\phi_2^\circ(\theta)}{q^\circ(\theta)\phi_1^\circ(\theta)} = \frac{\tilde{q}_2(\theta)}{\tilde{q}_1(\theta)} = \frac{\theta}{\bar{\zeta}}$$

is the same as in the case where investors can freely trade the risky assets, but the overall risk exposure

$$\|q^\circ(\theta)\phi^\circ(\theta)\| = \frac{1}{2a}\|\bar{\zeta}(\theta)\| = \frac{1}{2}\|q_0(\theta)\|$$

is smaller by a factor of two. The intuition for this result is that, since the loadings vector is determined only up to a multiplicative constant, offering the optimal fund $(1, \phi^\circ(\theta))$ to an investor of type θ is equivalent to offering him the fund $\frac{1}{a}\bar{\zeta}(\theta)$ that he would have picked on his own but with a fee equal $\frac{1}{2a}\|\bar{\zeta}(\theta)\|^2$ and, given this fee, the investor's optimal strategy is to invest half of what he would have on his own. Fourth, and last, the amount that an investor allocates to the fund (and hence the amount of fees that he pays) is increasing in both his risk tolerance and his type. As a result, the manager collects more fees from more optimistic or less risk averse investors, and his utility is increasing in $1/a$, $\bar{\zeta}$, and the parameter θ_H that determines the average belief of investors regarding the non-familiar asset.

B Proof of the results in Section 2

B.1 The revelation principle

Before proceeding with the proof of Proposition 1, we start by establishing some useful results about the investors' problem (1).

Lemma B.1 *The measure $q^* \in \mu_+(\mathcal{M})$ is optimal for an investor of type $\theta \in \Theta$ if and only*

if it satisfies

$$\int_{\mathcal{M}} \left\{ \phi(m)^\top \left(\zeta(\theta) - a \int_{\mathcal{M}} \phi(n) q^*(dn) \right) - \gamma(m) \right\} \nu(dm) \leq 0$$

for all measures ν in the set

$$\mathcal{F}(q^*) := \{ \nu \in \mu(\mathcal{M}) : \exists \beta > 0 \text{ such that } q^* + \beta \nu \in \mu_+(\mathcal{M}) \}.$$

In particular, the null measure is optimal for investors of type $\theta \in \Theta$ if and only if the menu is such that $\phi(m)^\top \zeta(\theta) \leq \gamma(m)$ for all $m \in \mathcal{M}$.

Proof. The first part follows from standard results on convex optimization in infinite dimensional spaces, see for example [Luenberger \(1969, Chapter 7\)](#). The second part follows from the first part by taking $p^* := 0$ and observing that $\mathcal{F}(0) = \mu_+(\mathcal{M})$. ■

Lemma B.2 *The value function of an investor of type $\theta \in \Theta$ satisfies*

$$v_i(\theta, \mathbf{m}) = \sup_{q \in \mu_{2,+}(\mathcal{M})} u(\theta, w_1(q, \mathbf{m}))$$

where $\mu_{2,+}(\mathcal{M})$ denotes the set of nonnegative measures on \mathcal{M} whose support consists in at most two distinct points.

Proof. A direct calculation shows that the optimization problem of an investor of type $\theta \in \Theta$ can be written as

$$v_i(\theta, \mathbf{m}) = \sup_{x \in \mathbf{R}^2} \sup_{q \in \mu_+^x(\mathcal{M})} \left\{ a \left(x_1 \zeta_1 + x_2 \theta - \int_{\mathcal{M}} \gamma(m) q(dm) \right) - \frac{a^2}{2} \|x\|^2 \right\}$$

where

$$\mu_+^x(\mathcal{M}) = \left\{ q \in \mu_+(\mathcal{M}) : \int_{\mathcal{M}} \phi(m) q(dm) = x \right\}.$$

By [Shapiro et al. \(2014, Proposition 6.40\)](#), we have that the inner supremum remains the same if one optimizes over $\mu_+^x(\mathcal{M}) \cap \mu_{2,+}(\mathcal{M})$ rather than over $\mu_+^x(\mathcal{M})$, and the result follows. ■

Proof of Proposition 1. Fix a menu $\mathbf{m}_0 = (\gamma_0, \phi_0, \mathcal{M}_0)$ and consider an alternative menu of the form $\mathbf{m} = (1, \phi, \Theta)$ for some fund loading function $\phi : \Theta \rightarrow \mathbf{R}^2$. As a first step we show that this fund loading function can be chosen in such a way that

the investment strategy $\pi(\theta, \phi(\theta))$ delivers each investor the same utility as his best response to \mathbf{m}_0 and generates the same amount of management fees.

By Lemma B.2 we have that given this menu each investor optimally allocates money to at most two funds. In order to construct the function $\phi(\theta)$, we therefore need to consider three mutually exclusive cases.

Case 0: If investors of type $\theta \in \Theta$ find it optimal to not invest in any of the proposed funds, then we know from Lemma B.1 that

$$\sup_{m \in \mathcal{M}_0} \left\{ \phi_0(m)^\top \xi(\theta) - \gamma_0(m) \right\} \leq 0.$$

Therefore, setting $\phi(\theta) = \frac{\phi_0(m)}{\gamma_0(m)}$ for some $m \in \mathcal{M}_0$ we get that $\pi(\theta, \phi(\theta)) = 0$ and the required properties follow.

Case 1: If the best response of investors of type $\theta \in \Theta$ is to allocate money to a single fund $m(\theta) \in \mathcal{M}_0$, then we have that

$$\begin{aligned} v_i(\theta, \mathbf{m}_0) &= u\left(\theta, rw_0 + q(\theta)R(\gamma_0(m(\theta)), \phi_0(m(\theta)))\right) \\ &= \frac{1}{2\|\phi_0(m(\theta))\|^2} \left(\phi_0(m(\theta))^\top \xi(\theta) - \gamma_0(m(\theta)) \right)^2, \end{aligned}$$

where

$$\begin{aligned} q(\theta) &= \arg \max_{q \in \mathbf{R}} \left\{ q \left(\phi_0(m(\theta))^\top \xi(\theta) - \gamma_0(m(\theta)) \right) - \frac{a}{2} q^2 \|\phi_0(m(\theta))\|^2 \right\} \\ &= \frac{1}{a\|\phi_0(m(\theta))\|^2} \left(\phi_0(m(\theta))^\top \xi(\theta) - \gamma_0(m(\theta)) \right). \end{aligned}$$

Setting $\phi(\theta) := \frac{\phi_0(m(\theta))}{\gamma_0(m(\theta))}$ shows that we have $\gamma_0(m(\theta))q(\theta) = \pi(\theta, \phi(\theta))$ and the desired properties now follow by observing that

$$\pi(\theta, \phi(\theta))R(1, \phi(\theta)) = q(\theta)R(\gamma_0(m(\theta)), \phi_0(m(\theta))).$$

Case 2: If the best response of investors of type $\theta \in \Theta$ is to allocate strictly positive

amounts to a pair of funds $(m_1(\theta), m_2(\theta)) \in \mathcal{M}_0$, then we have that

$$\begin{aligned} v_i(\theta, \mathbf{m}_0) &= u \left(\theta, rw_0 + \sum_{k=1}^2 q_k(\theta) R(\gamma_0(m_k(\theta)), \phi_0(m_k(\theta))) \right) \\ &= \frac{a^2}{2} \left\| \sum_{k=1}^2 q_k(\theta) \phi_0(m_k(\theta)) \right\|^2, \end{aligned}$$

where the vector

$$q(\theta) = \arg \max_{q \in \mathbf{R}^2} u \left(\theta, rw_0 + \sum_{k=1}^2 q_k R(\gamma_0(m_k(\theta)), \phi_0(m_k(\theta))) \right)$$

satisfies the first order conditions

$$\phi_0(m_k(\theta))^\top \xi(\theta) - \gamma_0(m_k(\theta)) = a \sum_{\ell=1}^2 q_\ell(\theta) \left(\phi_0(m_k(\theta))^\top \phi_0(m_\ell(\theta)) \right). \quad (34)$$

It follows that to satisfy the required properties for such types we need to choose the fund loading function in such a way that

$$\begin{aligned} \sum_{k=1}^2 \gamma_0(m_k(\theta)) q_k(\theta) &= \pi(\theta, \phi(\theta)), \\ \frac{a^2}{2} \left\| \sum_{k=1}^2 q_k(\theta) \phi_0(m_k(\theta)) \right\|^2 &= v_i(\theta, \mathbf{m}_0) = \frac{1}{2 \|\phi(\theta)\|^2} \left(\phi(\theta)^\top \xi(\theta) - 1 \right), \end{aligned}$$

and using the first order conditions (34) shows that the unique solution to this system is explicitly given by

$$\phi(\theta) = \frac{q_1(\theta) \phi_0(m_1(\theta)) + q_2(\theta) \phi_0(m_2(\theta))}{q_1(\theta) \gamma_0(m_1(\theta)) + q_2(\theta) \gamma_0(m_2(\theta))}.$$

To complete the proof, it now remains to show that the best response of an investor of type $\theta \in \Theta$ to the menu \mathbf{m} is indeed given by

$$q^*(\sigma; \theta, \mathbf{m}) = \mathbf{1}_{\{\theta \in \sigma\}} \pi(\theta, \phi(\theta)). \quad (35)$$

As is easily seen, every fund in the menu \mathbf{m} is a linear combination of funds in the menu \mathbf{m}_0 . Therefore, we have that

$$v_i(\theta, \mathbf{m}) \leq v_i(\theta, \mathbf{m}_0), \quad \theta \in \Theta$$

and the desired result follows by observing that the definition of the fund loading function $\phi(\theta)$ guarantees that $v_i(\theta, \mathbf{m}_0) = u(\theta, w_1(q^*, \mathbf{m}))$. \blacksquare

Proof of Proposition 2. Fix a menu $\mathbf{m} = (1, \phi, \Theta)$ and for each $\theta \in \Theta$ denote by \mathcal{F}_θ the set of feasible directions from the measure defined in (35). If the fund loading function is incentive compatible, then it follows from the result of Lemma B.1 that

$$\sup_{\nu \in \mathcal{F}_\theta} \int_{\Theta} \left\{ \phi(\theta')^\top \left(\xi(\theta) - a \int_{\Theta} \phi(\theta'') q^*(d\theta''; \theta, \mathbf{m}) \right) - 1 \right\} \nu(d\theta') \leq 0$$

for all $\theta \in \Theta$. Substituting the definition of the measure $q^*(\cdot; \theta, \mathbf{m})$ into the left hand side of this inequality, we obtain that

$$\sup_{\nu \in \mathcal{F}_\theta} \int_{\Theta} \left\{ \phi(\theta')^\top \xi(\theta) - 1 - \frac{\phi(\theta')^\top \phi(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \xi(\theta) - 1 \right) \right\} \nu(\theta') \leq 0$$

for all $\theta \in \Theta$, and the validity of (4) now follows by observing that the set \mathcal{F}_θ contains all nonnegative single point measures on Θ .

Conversely, assume that the fund loading function $\phi : \Theta \rightarrow \mathbf{R}^2$ is such that (4) holds, fix an arbitrary $\theta \in \Theta$, and let $\nu \in \mathcal{F}_\theta$ be a feasible direction from the measure defined in (35). By definition, we have that

$$\nu(\sigma) \geq -\mathbf{1}_{\{\theta \in \sigma\}} \pi(\theta, \phi(\theta)) / \beta$$

for some constant $\beta > 0$. Combining this property with (4) shows that

$$\begin{aligned} & \int_{\Theta} \left\{ \phi(\theta')^\top \left(\xi(\theta) - a \int_{\Theta} \phi(\theta'') q^*(d\theta''; \theta, \mathbf{m}) \right) - 1 \right\} \nu(d\theta') \\ &= \int_{\Theta} \left\{ \phi(\theta')^\top \xi(\theta) - 1 - \frac{\phi(\theta')^\top \phi(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \xi(\theta) - 1 \right)_+ \right\} \nu(d\theta') \\ &= \left\{ \phi(\theta)^\top \xi(\theta) - 1 - \left(\phi(\theta)^\top \xi(\theta) - 1 \right)_+ \right\} \nu(\{\theta\}) \\ &+ \int_{\Theta \setminus \{\theta\}} \left\{ \phi(\theta')^\top \xi(\theta) - 1 - \frac{\phi(\theta')^\top \phi(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \xi(\theta) - 1 \right)_+ \right\} \nu(d\theta') \\ &\leq - \left(1 - \phi(\theta)^\top \xi(\theta) \right)_+ \nu(\{\theta\}) \\ &\leq \left(1 - \phi(\theta)^\top \xi(\theta) \right)_+ \frac{\left(\phi(\theta)^\top \xi(\theta) - 1 \right)_+}{\beta \|\phi(\theta)\|^2} = 0 \end{aligned}$$

and the incentive compatibility of the fund loading function now follows from Lemma

B.1 and the arbitrariness of the pair $(\theta, \nu) \in \Theta \times \mathcal{F}_\theta$. ■

B.2 The relaxed problem

Proof of Lemma 1. Let $\phi \in \Phi$ be an incentive compatible fund loading function, and assume that the set

$$A := \{\theta \in \Theta : \pi(\theta, \phi(\theta)) > 0\} = \{\theta \in \Theta : \phi(\theta)^\top \zeta(\theta) - 1 > 0\}$$

is non empty, for otherwise the statement is trivial. Since $\Phi \subseteq AC(\Theta; \mathbf{R}^2)$ and

$$\|\phi(\theta)\| \|\zeta(\theta_H)\| \geq \|\phi(\theta)\| \|\zeta(\theta)\| \geq \phi(\theta)^\top \zeta > 1$$

on the set A , we have that $v \in AC(\Theta; \mathbf{R})$. This implies that $v(\theta)$ is differentiable at almost every $\theta \in \Theta$ and a direct calculation shows that we have

$$\begin{aligned} \dot{v}(\theta) = \frac{1}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \zeta(\theta) - 1 \right) \left(\dot{\phi}(\theta)^\top \zeta(\theta) + \phi_2(\theta) \right) \\ - \frac{\dot{\phi}(\theta)^\top \phi(\theta)}{\|\phi(\theta)\|^4} \left(\phi(\theta)^\top \zeta(\theta) - 1 \right)^2 \end{aligned} \quad (36)$$

for almost every $\theta \in A$ and $\dot{v}(\theta) = 0$ for all $\theta \in A^c$. Equation (4) implies that for every $\theta \in A$ the absolutely continuous function

$$\theta' \mapsto F_\theta(\theta') = \left(\phi(\theta')^\top \zeta(\theta) - 1 \right) - \frac{\dot{\phi}(\theta')^\top \phi(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \zeta(\theta) - 1 \right)$$

attains a maximum at the point $\theta' = \theta$. In particular, we have that

$$\left. \frac{\partial F_\theta(\theta')}{\partial \theta'} \right|_{\theta'=\theta} = \dot{\phi}(\theta)^\top \zeta(\theta) - 1 - \frac{\dot{\phi}(\theta)^\top \phi(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \zeta(\theta) - 1 \right) = 0$$

for almost every $\theta \in A$. Substituting this expression into (36) shows that

$$\dot{v}(\theta) = \frac{\phi_2(\theta)}{\|\phi(\theta)\|^2} \left(\phi(\theta)^\top \zeta(\theta) - 1 \right), \quad \text{a.e. } \theta \in A, \quad (37)$$

and combining this identity with the definition of $v(\theta)$ we obtain that

$$2v(\theta) \|\phi(\theta)\|^2 = \left(\phi(\theta)^\top \zeta(\theta) - 1 \right)^2, \quad (38)$$

$$\dot{v}(\theta) \|\phi(\theta)\|^2 = \phi_2(\theta) \left(\phi(\theta)^\top \zeta(\theta) - 1 \right), \quad (39)$$

for almost every $\theta \in A$. Solving this system gives

$$\phi_{\pm}(\theta) = \frac{\left(\sqrt{2v(\theta) - [\dot{v}(\theta)]^2}, \dot{v}(\theta)\right)^{\top}}{\theta\dot{v}(\theta) - 2v(\theta) \pm \xi\sqrt{2v(\theta) - [\dot{v}(\theta)]^2}}, \quad \text{a.e. } \theta \in A, \quad (40)$$

for almost every $\theta \in A$ and we claim that only $\phi_+(\theta)$ is consistent with the definition of the set A . Indeed, letting

$$p_{\pm}(\theta) := a\pi(\theta, \phi_{\pm}(\theta)) = \frac{\phi_{\pm}(\theta)^{\top} \zeta(\theta) - 1}{\|\phi_{\pm}(\theta)\|^2}, \quad (41)$$

and using (37) in conjunction with (40) and the definition of $v(\theta)$ we obtain that

$$\begin{aligned} p_{\pm}(\theta) &= x\dot{v}(\theta) - 2v(\theta) \pm \xi\sqrt{2v(\theta) - [\dot{v}(\theta)]^2} \\ &= x\phi_{\pm}(\theta)p_{\pm}(\theta) - \|\phi_{\pm}(\theta)\|^2 p_{\pm}(\theta)^2 \pm \xi\sqrt{\|\phi_{\pm}(\theta)\|^2 p_{\pm}(\theta)^2 - \phi_{2,\pm}(\theta)^2 p(\theta)^2} \\ &= x\phi_{\pm}(\theta)p_{\pm}(\theta) - \|\phi_{\pm}(\theta)\|^2 p_{\pm}(\theta)^2 \pm p_{\pm}(\theta)\phi_{1,\pm}(\theta)\xi \\ &= p_{\pm}(\theta) - (1 \mp 1)p_{\pm}(\theta)\phi_{1,\pm}(\theta)\xi \end{aligned} \quad (42)$$

for almost every $\theta \in A$. This shows that

$$\mathbf{1}_{\{\theta \in A\}} \left(\phi(\theta)^{\top} - \frac{\left(\sqrt{2v(\theta) - [\dot{v}(\theta)]^2}, \dot{v}(\theta)\right)}{\theta\dot{v}(\theta) - 2v(\theta) + \xi\sqrt{2v(\theta) - [\dot{v}(\theta)]^2}} \right) = 0,$$

and it now remains to establish the validity of (7) and (8). On the set A these properties follow from (41), (42), and the fact that we have

$$2v(\theta) - [\dot{v}(\theta)]^2 = 2\phi_1(\theta)^2 v(\theta)$$

as a result of (38) and (39). On the set A^c we have that

$$\dot{v}(\theta) = v(\theta) = \frac{1}{\|\phi(\theta)\|^2} \left(\phi(\theta)^{\top} \zeta(\theta) - 1 \right)_+ = 0,$$

and the desired result follows by observing that the function $F(\theta, v(\theta), \dot{v}(\theta))$ is such that $F(\theta, 0, 0) = 0$ for all $\theta \in \Theta$. ■

The proof of Proposition 3 will be carried out through a series of lemmas. The first lemma establishes the uniqueness of the solution and shows that it attains the supremum in the relaxed problem.

Lemma B.3 Assume that $v^* \in C^2(\Theta; \mathbf{R})$ is a solution to the boundary value problem defined by (10), (11), and (12). Then, we have

$$V = \int_{\Theta} F(\theta, v^*(\theta), \dot{v}^*(\theta)) d\theta.$$

In particular, there can be at most one classical solution to the boundary value problem.

Proof. Let $v^* : \Theta \rightarrow \mathbf{R}$ be as in the statement. Since $v^* \in C^2(\Theta; \mathbf{R})$ by assumption we necessarily have that $v^* \in \mathcal{V}$ for otherwise

$$1 + \dot{v}^*(\theta) = \frac{1}{v^*(\theta)} \left([\dot{v}^*(\theta)]^2 + \frac{3}{2\xi} \left(2v^*(\theta) - [\dot{v}^*(\theta)]^2 \right)^{\frac{3}{2}} \right)$$

would not be real valued on Θ . Let now $v \in \mathcal{V}$ be another feasible function. Using the result of Lemma B.4 below, we deduce that

$$\begin{aligned} \int_{\Theta} \left(F(x, v(\theta), \dot{v}(\theta)) - F(x, v^*(\theta), \dot{v}^*(\theta)) \right) d\theta &\leq \Delta(v, v^*) \\ &:= \int_{\Theta} \left((v(\theta) - v^*(\theta)) F_{v^*(\theta)}^*(\theta) + (\dot{v}(\theta) - \dot{v}^*(\theta)) F_{\dot{v}^*(\theta)}^*(\theta) \right) d\theta, \end{aligned}$$

where we have set

$$F_k^*(\theta) := F_k(x, v^*(\theta), \dot{v}^*(\theta)), \quad k \in \{v^*(\theta), \dot{v}^*(\theta)\}.$$

Now, since $v^* \in C^2(\Theta; \mathbf{R})$ by assumption we have that $\dot{v}^* \in C^1(\Theta; \mathbf{R})$, and we may thus integrate by parts to show that

$$\begin{aligned} \Delta(v, v^*) &= \left((v - v^*)(\theta) F_p^*(\theta) \right) \Big|_{\theta=0}^{\theta_H} + \int_{\Theta} (v - v^*)(\theta) \left(F_v^*(\theta) - \frac{d}{d\theta} F_p^*(\theta) \right) d\theta \\ &= (v(\theta_H) - v^*(\theta_H)) F_p^*(\theta_H) - (v(0) - v^*(0)) F_p^*(0) = 0, \end{aligned}$$

where the last two equalities follow from the fact that v^* solves the Euler-Lagrange equation (9) subject to (11). To complete the proof assume that $(v_i)_{i=1}^2 \in C^2(\Theta; \mathbf{R})$ are distinct solutions, and let

$$v^*(\theta) = \frac{1}{2} \sum_{i=1}^2 v_i(\theta). \tag{43}$$

By the first part of the proof we have that

$$V = \int_{\Theta} F(\theta, v_i(\theta), \dot{v}_i(\theta)) d\theta, \quad i \in \{1, 2\},$$

and combining this identity with (43) and Lemma B.4 we deduce that

$$V = \frac{1}{2} \sum_{i=1}^2 \int_{\Theta} F(\theta, v_i(\theta), \dot{v}_i(\theta)) d\theta < \int_{\Theta} F(x, v^*(\theta), \dot{v}^*(\theta)) d\theta.$$

Since $v^* \in \mathcal{V}$ this inequality contradicts the fact that the functions $(v_i)_{i=1}^2$ both attain the supremum in (\mathcal{R}) , and establishes the required uniqueness. ■

Lemma B.4 *Let*

$$\mathcal{O} := \left\{ (v, p) \in \mathbf{R}^2 : v \neq 0 \text{ and } 2v - p^2 \geq 0 \right\}.$$

The function $F(\theta, y, p)$ is strictly concave in $(v, p) \in \mathcal{O}$ for any fixed $\theta \in \Theta$.

Proof. A direct calculation shows that

$$\frac{\partial^2 F}{\partial v \partial p}(\theta, v, p) = \zeta (2v - p^2)^{-\frac{3}{2}} \begin{bmatrix} -1 & p \\ p & -2v \end{bmatrix}.$$

The determinant and trace of this matrix are, respectively, strictly positive and strictly negative for all $(v, p) \in \mathcal{O}$. Therefore, its eigenvalues are strictly negative. ■

To prove the existence of a solution to the boundary value problem (10)–(12) we start by showing that for any initial condition q in an appropriate interval the initial value problem given by (10) subject to

$$\dot{v}(0) = v(0) - q = 0$$

admits a unique classical solution on the positive real line. Then, we show that the initial condition q can be chosen in such a way as to satisfy the boundary condition (12) at the upper end point of the type space.

Lemma B.5 *The initial value problem*

$$v(\theta) (1 + \dot{v}(\theta)) - [\dot{v}(\theta)]^2 = \frac{3}{2\zeta} \left(2v(\theta) - [\dot{v}(\theta)]^2 \right)^{\frac{3}{2}}, \quad (44)$$

$$\dot{v}(0) = v(0) - q = 0, \quad (45)$$

admits a unique solution $v(\theta) = v(\theta; q)$ in $C^2(\mathbf{R}_+; \mathbf{R})$ for any $q > \frac{1}{18}\xi^2$. This solution is decreasing in q as well as strictly increasing and strictly convex in θ with

$$\inf_{\theta \geq 0} \left(\mathbf{1}_{\{q \in \mathcal{C}_1\}} - \mathbf{1}_{\{q \in \mathcal{C}_2\}} \right) (1 - \ddot{v}(\theta)) \geq 0,$$

where we have set $\mathcal{C}_1 := \left(\frac{1}{18}\xi^2, \frac{2}{9}\xi^2 \right]$ and $\mathcal{C}_2 := (2\xi^2/9, \infty)$.

Proof. Let q be fixed and write the initial value problem defined by (44) and (45) as a system of first order differential equations

$$0 = X'(\theta) - G(X(\theta)) = X(0) - (q, 0)^\top, \quad \theta \geq 0 \quad (46)$$

with the function

$$G(X) := \left(X_2, 1 + \frac{1}{\xi} \left(1 - \frac{X_2^2}{2X_1} \right) \left(3\sqrt{2X_1 - X_2^2} - 2\xi \right) \right)^\top.$$

Since $G \in C^1(\mathcal{O}; \mathbf{R}^2)$ it follows from [Hirsch et al. \(2013, p.387\)](#) that the initial value problem (46) admits a unique solution that is defined on $[0, \bar{\theta})$ for some $\bar{\theta} \leq \infty$. Before showing that this solution is actually defined on the whole positive real line, we start by establishing the other properties listed in the statement.

Letting $\theta \rightarrow 0$ in the differential equation and using the fact that $\dot{v}(0) = 0$ shows that we have

$$\xi(1 - \ddot{v}(0)) = 2\xi - \sqrt{18q}. \quad (47)$$

To proceed further, we distinguish two cases. If $q \in \mathcal{C}_2$, then $\ddot{v}(0) > 1$, and we claim that the second derivative may reach one but never goes below. To see this, consider the function defined by

$$b(\theta) := 2v(\theta) - [\dot{v}(\theta)]^2,$$

and assume that the solution is such that

$$\hat{\theta} := \inf \{ \theta \in [0, \bar{\theta}) : \dot{v}(\theta) = 1 \} < \bar{\theta},$$

for otherwise there is nothing to prove. Evaluating the differential equation (44) at the

point $\hat{\theta}$ shows that

$$b(\hat{\theta}) \left(2\bar{\zeta} - 3\sqrt{b(\hat{\theta})} \right) = 0$$

and it follows that either $b(\hat{\theta}) = 0$ or $b(\hat{\theta}) = 4\bar{\zeta}^2/9$. Since $\ddot{v}(\theta) > 1$ on $[0, \hat{\theta})$ we have that $b(\theta)$ is strictly decreasing on that interval, and using this property in conjunction with the fact that $b(0) = 2q > 4\bar{\zeta}^2/9$, we deduce that $b(\hat{\theta}) = 4\bar{\zeta}^2/9$. This in turn implies that $v(\theta)$ solves the initial value problem

$$\begin{aligned} w(\theta) (1 + \ddot{w}(\theta)) &= [\dot{w}(\theta)]^2 + \frac{3}{2\bar{\zeta}} \left(2w(\theta) - [\dot{w}(\theta)]^2 \right)^{\frac{3}{2}}, \\ w(\hat{\theta}) &= v(\hat{\theta}), \\ \dot{w}(\hat{\theta}) &= \left(2v(\hat{\theta}) - 4\bar{\zeta}^2/9 \right)^{\frac{1}{2}}, \end{aligned}$$

on the interval $[\hat{\theta}, \bar{\theta})$ and, since the unique solution to this initial value problem is explicitly given by

$$w(\theta) = v(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^2 + (\theta - \hat{\theta}) \left(2v(\hat{\theta}) - 4\bar{\zeta}^2/9 \right)^{\frac{1}{2}},$$

we conclude that $\ddot{v}(\theta) = \ddot{w}(\theta) = 1$ for all $\theta \in [\hat{\theta}, \bar{\theta})$. Because $\ddot{v}(\theta) \geq 1$ for all $\theta \in [0, \bar{\theta})$ we have that the solution is strictly convex and combining this property with the fact that $\dot{v}(0) = 0$ we deduce that it is strictly increasing.

Assume next that $q \in \mathcal{C}_1$ so that $\ddot{v}(0) \in (0, 1]$. If $q = \frac{2}{9}\bar{\zeta}^2$ then a direct calculation shows that the unique solution to the initial value problem is

$$v(\theta) = \frac{2}{9}\bar{\zeta}^2 + \frac{1}{2}\theta^2,$$

and it follows that we have $\ddot{v}(\theta) = 1$ for all $\theta \in [0, \bar{\theta})$. Now assume that the initial condition $q < \frac{2}{9}\bar{\zeta}^2$, and denote by

$$\theta_0 := \inf \{ \theta \in [0, \bar{\theta}) : \ddot{v}(\theta) \geq 1 \}$$

the first point at which the second derivative reaches one. Since $\ddot{v}(0) < 1$ we have that $\theta_0 > 0$. Assume that $\theta_0 < \bar{\theta}$. Since the solution is twice continuously differentiable on its domain of definition we have that $\ddot{v}(\theta_0) = 1$, and it thus follows from (44) that the

function $\ddot{v}(\theta)$ solves the differential equation

$$\dot{w}(\theta) = L(x, w(\theta)) := (1 - w(\theta)) \left(9\sqrt{b(\theta)} - 2\zeta \right) \frac{\dot{v}(\theta)}{v(\theta)}, \quad (48)$$

$$w(\theta_1) = 1. \quad (49)$$

Since the functions $v(\theta)$ and $\dot{v}(\theta)$ are continuous in a neighbourhood of θ_1 we have that $L(x, w)$ is continuously differentiable in a neighbourhood of $(\theta_1, 1)$ and it follows that there exists an $\varepsilon > 0$ such that (48)–(49) admits a unique solution in $(\theta_1 - \varepsilon, \theta_1 + \varepsilon)$. Because $\ddot{v}(\theta)$ and the constant function $w(\theta) \equiv 1$ are both solutions to that differential equation, uniqueness implies that we must have $\ddot{v}(\theta) = 1$ in a left neighbourhood of the point θ_1 . This contradicts the definition of θ_1 as the first point where the second derivative reaches one, and thus shows that we have $\theta_1 = \infty$.

Since $\dot{v}(0) = 0$ and $\ddot{v}(0) > 0$ the strict increase of the solution will follow from its strict convexity. Assume towards a contradiction that the solution is not strictly convex on its domain of definition so that

$$\theta_0 := \inf \{ \theta \in [0, \bar{\theta}) : \ddot{v}(\theta) \leq 0 \} < \bar{\theta}$$

By continuity we have that

$$0 = \ddot{v}(\theta_0) < \ddot{v}(\theta), \quad \theta \in [0, \theta_0), \quad (50)$$

and it follows that $v(\theta)$ and $\dot{v}(\theta)$ are strictly increasing on $[0, \theta_0)$. Differentiating both sides of (44) shows that

$$\ddot{v}(\theta) = (1 - \dot{v}(\theta)) \left(9\sqrt{b(\theta)} - 2\zeta \right) \frac{\dot{v}(\theta)}{v(\theta)} \geq 0, \quad (51)$$

where the inequality follows from the nonnegativity and increase of $v(\theta)$, and the fact that $\dot{v}(\theta) \leq 1$, which implies that

$$9\sqrt{b(\theta)} - 2\zeta \geq 9\sqrt{b(0)} - 2\zeta = 9\sqrt{2q} - 2\zeta \geq \zeta, \quad \theta \in [0, \bar{\theta}).$$

This shows that the function $\ddot{v}(\theta)$ is increasing on the interval $[0, \theta_0)$ and implies that we have $\ddot{v}(\theta_0) \geq \ddot{v}(0) > 0$, which contradicts equation (50).

To complete the first part of the proof it now remains to show that the solution is defined on the whole positive real line. Standard results on first order differential systems (see e.g. [Hirsch et al. \(2013, p.398\)](#)) imply that this can fail to be the case only

if the solution becomes unbounded or reaches the boundary of \mathcal{O} . Assume first that $q \in \mathcal{C}_1$. Using the fundamental theorem of calculus in conjunction with the fact that the solution is non decreasing and such that $\ddot{v}(\theta) \leq 1$ gives

$$0 \leq \dot{v}(\theta) = \dot{v}(0) + \int_0^\theta \ddot{v}(\theta) dx \leq \dot{v}(0) + \theta = \theta$$

and, therefore,

$$q = v(0) \leq v(\theta) = q + \int_0^\theta \dot{v}(\theta) dx = q + \int_0^\theta x dx = q + \frac{1}{2}\theta^2,$$

which shows that the solution cannot grow unbounded. On the other hand, because the solution is such that $\ddot{v}(\theta) \leq 1$ we have that the function $b(\theta)$ is nondecreasing, and it follows that

$$b(\theta) = 2v(\theta) - [\dot{v}(\theta)]^2 \geq b(0) = 2v(0) = 2q > 0,$$

which shows that the solution remains in the interior of \mathcal{O} . Assume next that the initial condition $q \in \mathcal{C}_2$ and consider the function $b(\theta)$. Since $q \in \mathcal{C}_2$ we know from the first part of the proof that this function is decreasing and such that $b(0) > 4\zeta^2/9$. If the function $b(\theta)$ remains above $q^* = 4\zeta^2/9$, then we have that the solution never reaches the boundary of the set \mathcal{O} . On the other hand, if the function $b(\theta)$ reaches q^* at some point $\theta^* \in [0, \bar{\theta})$, then it follows from (44) that $\ddot{v}(\theta^*) = 1$, and the same arguments as in the first part of the proof then show that

$$v(\theta) = v(\theta^*) + \frac{1}{2}(\theta - \theta^*)^2 + (\theta - \theta^*) \left(2v(\theta^*) - 4\zeta^2/9 \right)^{\frac{1}{2}}$$

for all $\theta \in [\theta^*, \bar{\theta})$. This in turn implies that $b(\theta) = 4\zeta^2/9$ for all $\theta \in [\theta^*, \bar{\theta})$, and it follows that the solution to the initial value problem never reaches the boundary of the set \mathcal{O} . Finally, differentiating (44) shows that

$$\ddot{v}(\theta) = (1 - \ddot{v}(\theta)) \left(9\sqrt{b(\theta)} - 2\zeta \right) \frac{\dot{v}(\theta)}{v(\theta)} \leq 0,$$

where the inequality follows from the fact that $\ddot{v}(\theta) \geq 1$ and $b(\theta) \geq 4\zeta^2/9$. This implies that $\ddot{v}(\theta)$ is decreasing, and combining this property with the strict increase of the solution we obtain that

$$v(0) \leq v(\theta) = q + \int_0^\theta (\theta - x)\ddot{v}(\theta) dx \leq q + \frac{1}{2}\ddot{v}(0)\theta^2,$$

and it follows that the solution cannot grow unbounded.

To complete the proof it now remains to show that the solution is decreasing in the initial condition. Since the right hand side of (46) belongs to $C^1(\mathcal{O}; \mathbf{R})$ we know from [Hirsch et al. \(2013, p.395\)](#) that the corresponding flow is continuous and the desired result will follow from the Kamke-Müller theorem (see, e.g., [Müller \(1927\)](#)) provided that the Jacobian matrix

$$J(\theta) := \nabla G(X(\theta)) = \left[\frac{\partial G_i}{\partial X_j}(X(\theta)) \right]_{i,j=1}^2$$

is of Metzler type for all $\theta \geq 0$. A direct calculation shows that the off-diagonal terms of this matrix are explicitly given by $J_{12}(\theta) = 1$ and

$$J_{21}(\theta) = \frac{3(v(\theta) + [\dot{v}(\theta)]^2) \sqrt{2v(\theta) - [\dot{v}(\theta)]^2} - 2\bar{\zeta}[\dot{v}(\theta)]^2}{2\bar{\zeta}v(\theta)^2}.$$

Assume towards a contradiction that this function is not positive throughout the type space. Since $J_{21}(0) > 0$ this implies that there exists $\bar{\theta} > 0$ such that

$$J_{21}(\bar{\theta}) < 0 = J_{21}(\bar{\theta}). \quad (52)$$

The assumption that $J_{21}(\bar{\theta}) = 0$ implies that we have

$$\bar{\zeta} = \frac{3}{2[\dot{v}(\bar{\theta})]^2} \left(v(\bar{\theta}) + [\dot{v}(\bar{\theta})]^2 \right) \sqrt{2v(\bar{\theta}) - [\dot{v}(\bar{\theta})]^2}.$$

Using this expression in conjunction with the fact that the function $v(\theta)$ solves (44) the shows that we have

$$J_{21}(\bar{\theta}) = \frac{6\sqrt{2v(\bar{\theta}) - [\dot{v}(\bar{\theta})]^2}}{\dot{v}(\bar{\theta}) (v(\bar{\theta}) + [\dot{v}(\bar{\theta})]^2)} \left\{ \left(v(\bar{\theta}) - [\dot{v}(\bar{\theta})]^2 \right)^2 + v(\bar{\theta})[\dot{v}(\bar{\theta})]^2 \right\} \geq 0$$

which contradicts (52). ■

Lemma B.6 *For any $\theta > 0$, there exists a unique $q = q(\theta) \in \mathcal{C}_1$ such that*

$$\Lambda(\theta, q) := \theta - \frac{\bar{\zeta}\dot{v}(\theta; q)}{\sqrt{2v(\theta; q) - [\dot{v}(\theta; q)]^2}} = 0, \quad (53)$$

where $v(\theta; q)$ denotes the unique solution to (44) and (45). Furthermore, the function $q(\theta)$ is continuous, strictly decreasing, and such that $\lim_{\theta \rightarrow 0} q(\theta) = \frac{1}{8}\bar{\zeta}^2$.

Proof. Fix an arbitrary $\theta_H > 0$. Since the right hand side of (46) belongs to $C^1(\mathcal{O}; \mathbf{R})$, we know from Hirsch et al. (2013, p.395) that the corresponding flow is continuous, and it follows that the function $q \mapsto \Lambda(\theta; q)$ is continuous on \mathcal{C}_1 . A direct calculation shows that for $q \in \partial\mathcal{C}_1$ this flow is explicitly given by

$$\begin{aligned} X\left(\theta; \frac{1}{18}\xi^2\right)^\top &= \left(\frac{1}{18}\xi^2, 0\right), \\ X\left(\theta; \frac{2}{9}\xi^2\right)^\top &= \left(\frac{2}{9}\xi^2 + \frac{1}{2}\theta^2, \theta\right). \end{aligned}$$

Substituting these expressions in (53) then shows that

$$\Lambda\left(\theta, \frac{2}{9}\xi^2\right) = -\frac{1}{2}\theta < 0 < \theta = \Lambda\left(\theta, \frac{1}{18}\xi^2\right) \quad (54)$$

and the existence of a solution now follows from the intermediate value theorem. To complete the first part of the proof assume that there exists $\theta_H > 0$ such that (53) admits two solutions $q_1 \neq q_2$. Then the functions $v(\theta; q_1)$ and $v(\theta; q_2)$ both solve the boundary value problem associated with the upper end point $\theta_H > 0$ but differ in a right neighbourhood of the origin. This contradicts the conclusion of Lemma B.3 and establishes the required uniqueness.

According to the first part of the proof we have that the solution mapping defines a function $q : (0, \infty) \rightarrow \mathcal{C}$ and, since $v(\theta; q)$ and $\dot{v}(\theta; q)$ are both continuous, we have that this function is continuous. Indeed, if this was not the case, then there would exist a point $\theta_0 > 0$ and two sequences $(\theta_{i,n})_{n=1}^\infty \subseteq (0, \infty)$ such that

$$\theta_0 = \lim_{n \rightarrow \infty} \theta_{1,n} = \lim_{n \rightarrow \infty} \theta_{2,n},$$

and

$$q_1 := \lim_{n \rightarrow \infty} q(\theta_{1,n}) \neq \lim_{n \rightarrow \infty} q(\theta_{2,n}) := q_2.$$

The continuity of the functions $(v(\theta; q), \dot{v}(\theta; q))$ and the definition of $q(\cdot)$ then imply that we have

$$0 = \lim_{n \rightarrow \infty} |\Lambda(\theta_{i,n}, q(\theta_{i,n}))| = |\Lambda(\theta_0, q_i)|,$$

and it follows that $q_i \in \text{int}(\mathcal{C}_1)$ for otherwise (54) would imply that the term on the right hand side is strictly positive. This contradicts the fact that the solution to (53) is unique in \mathcal{C}_1 for every $\theta > 0$, and establishes the required continuity.

Now assume that the solution mapping is not strictly monotone. By continuity this implies that there exist $\theta_1, \theta_2 > 0$ such that $\theta_1 \neq \theta_2$ and $q(\theta_1) = q(\theta_2) := q^* \in \mathcal{C}_1$. The definition of the solution mapping then implies that

$$\Lambda(\theta_1, q^*) = \Lambda(\theta_2, q^*),$$

which contradicts the conclusion of Lemma B.7 below. Next, we claim that the solution mapping is such that $q(\theta) < \frac{1}{8}\zeta^2$ for all $\theta > 0$. Indeed, if we had that $q(\theta_0) \geq \frac{1}{8}\zeta^2$ for some $\theta_0 > 0$ then the equation

$$\theta \mapsto \Lambda(\theta, (q(\theta_0))) = 0$$

would admit a strictly positive solution given by $\theta = \theta_0$, and this would contradict the conclusion of Lemma B.7 below.

The above results show that the solution mapping is continuous, monotone and bounded on $(0, \infty)$. Therefore, the limit $q(0) := \lim_{x \rightarrow 0} q(\theta)$ exists and is finite, and the proof will be complete once we show that this limit equals $\frac{1}{8}\zeta^2$. Since $\dot{v}(0; q) = 0$ for all $q \in \mathcal{C}_1$ we have that

$$\lim_{\theta \rightarrow 0} \frac{\dot{v}(\theta, q(\theta))}{\theta} = \ddot{v}(0; q(0)).$$

Using this identity in conjunction with (47) and the definition of the solution mapping we obtain that

$$0 = \lim_{\theta \rightarrow 0} \frac{\Lambda(\theta, q(\theta))}{\theta} = 1 - \frac{\zeta \ddot{v}(0; q(0))}{\sqrt{2q(0)}} = 1 + \frac{\zeta}{\sqrt{2q(0)}} \left(1 - \frac{\sqrt{18q(0)}}{\zeta} \right),$$

and solving for $q(0)$ gives the desired result. Knowing that the solution mapping is strictly monotone and such that $q(\theta) \leq q(0)$ for all $\theta \geq 0$, we then deduce that it is strictly decreasing and the proof is complete. ■

Lemma B.7 For $q \in \mathcal{C}_1$ the equation $\Lambda(\theta, q) = 0$ admits a solution $\theta > 0$ only if $q \leq \frac{1}{8}\zeta^2$, and in this case there is at most one solution.

Proof. A direct calculation using (44) shows that

$$\frac{\partial \Lambda}{\partial \theta}(\theta, q) = \frac{\zeta}{\sqrt{2v(\theta; q) - [\dot{v}(\theta; q)]^2}} - 2.$$

Since $\dot{v}(\theta; q) \leq 1$ for all $(\theta, q) \in \mathbf{R}_+ \times \mathcal{C}_1$, by Lemma B.5 we have that this derivative

is non increasing as a function of θ , and it follows that the function $\theta \mapsto \Lambda(\theta, q)$ is concave. If $q > \frac{1}{8}\xi^2$ then this concavity implies that

$$\frac{\partial \Lambda}{\partial \theta}(\theta, q) \leq \frac{\partial \Lambda}{\partial \theta}(0, q) = \frac{\xi}{\sqrt{2q}} - 2 < 0,$$

and it follows that the only solution to $\Lambda(\theta, q) = 0$ is given by $\theta = 0$. On the other hand, if $q \leq \frac{1}{8}\xi^2$ then $\frac{\partial \Lambda}{\partial \theta}(0, q) \geq 0$, and the concavity of the function $\theta \mapsto \Lambda(\theta, q)$ implies that there can be at most one $\theta > 0$ such that $\Lambda(\theta, q) = 0$. ■

Proof of Proposition 3. By construction we have that the function $v^*(\theta) = v(\theta; q(\theta_H))$ belongs to $C^2(\Theta; \mathbf{R})$, and solves the boundary value problem. Therefore, it follows from the result of Lemma B.3 that this function is the unique such solution and that it attains the supremum in the relaxed problem. Furthermore, since $\theta_H > 0$, we know from Lemma B.6 that $q(\theta_H) > \frac{1}{8}\xi^2$, and it thus follows from Lemma B.5 that the function $v^*(\theta)$ is strictly increasing and strictly convex on Θ . ■

B.3 The optimal fund menu

Proof of Theorem 1. Let us start by establishing (13). Since $v^*(0) = q(\theta_H) \in \mathcal{C}_1$ by Lemma B.6 we know from Lemma B.5 that

$$\inf_{\theta \in \Theta} (1 - \ddot{v}^*(\theta)) \geq 0 \tag{55}$$

Using this property in conjunction with the fundamental theorem of calculus then shows that we have

$$2v^*(\theta) - [\dot{v}^*(\theta)]^2 = 2q(\theta_H) + \int_0^\theta 2\dot{v}^*(x) (1 - \ddot{v}^*(x)) dx > \frac{1}{9}\xi^2. \tag{56}$$

As a result, (13) will follow once we show that

$$c(\theta) := \frac{aF(\theta, v^*(\theta), \dot{v}^*(\theta))}{\sqrt{2v^*(\theta) - [\dot{v}^*(\theta)]^2}} = \xi + \frac{\theta \dot{v}^*(\theta) - 2v^*(\theta)}{\sqrt{2v^*(\theta) - [\dot{v}^*(\theta)]^2}} > 0 \tag{57}$$

for all $\theta \in \Theta$. Since $v^*(0) = q(\theta_H) \in \mathcal{C}_1$ by Lemma B.6 we have that $c(0) \geq \xi/3$ is strictly positive, and it is therefore sufficient to show that $c(\theta)$ is non decreasing. Differentiating (57) we find that

$$\dot{c}(\theta) = \frac{(\theta - \dot{v}^*(\theta)) (2v^*(\theta)\ddot{v}^*(\theta) - [\dot{v}^*(\theta)]^2)}{(2v^*(\theta) - [\dot{v}^*(\theta)]^2)^{\frac{3}{2}}}.$$

Using (55) we deduce that

$$\theta - \dot{v}^*(\theta) = \int_0^\theta (1 - \ddot{v}^*(x)) dx \geq 0.$$

On the other hand, using the fact that $v^*(0) = q(\theta_H) \in \mathcal{C}_1$ in conjunction with the fundamental theorem of calculus, (47), and (51) we obtain that

$$\begin{aligned} \ell(\theta) &:= 2v^*(\theta)\ddot{v}^*(\theta) - [\dot{v}^*(\theta)]^2 = \ell(0) + \int_0^\theta 2v^*(x)\ddot{v}^*(x) dx \\ &\geq \ell(0) = 2v^*(0) \left(\sqrt{\frac{18v^*(0)}{\xi^2}} - 1 \right) \geq 0, \end{aligned} \quad (58)$$

and the desired result now follows from (56).

Let us now turn to the second part of the statement. Since $v^* \in C^2(\Theta; \mathbf{R})$ we have that $\phi^* \in C^1(\Theta; \mathbf{R}^2)$, and the feasibility of ϕ^* will follow once we show that

$$L(\theta, \theta') := \left(\phi^*(\theta')^\top \xi(\theta) - 1 \right) - \frac{\phi^*(\theta')^\top \phi^*(\theta)}{\|\phi^*(\theta)\|^2} \left(\phi^*(\theta)^\top \xi(\theta) - 1 \right)_+$$

is non positive for all $(\theta, \theta') \in \Theta^2$. Substituting the definition of $\phi^*(\theta)$ into the left hand side, and using the fact that

$$\frac{1}{\|\phi^*(\theta)\|^2} \left(\phi^*(\theta)^\top \xi(\theta) - 1 \right)_+ = aF(\theta, v^*(\theta), \dot{v}^*(\theta)) > 0,$$

as a result of (13) we obtain that

$$L(\theta, \theta') = - \frac{2v^*(\theta)(g(\theta)g(\theta') - 2v^*(\theta') + (\theta' - \theta + \dot{v}^*(\theta))\dot{v}^*(\theta'))}{a^3F(\theta', v^*(\theta'), \dot{v}^*(\theta'))F(\theta, v^*(\theta), \dot{v}^*(\theta))^2}$$

with

$$g^*(\theta) := \sqrt{2v^*(\theta) - [\dot{v}^*(\theta)]^2}.$$

Since the functions $v^*(\theta)$ and $F(\theta, v^*(\theta), \dot{v}^*(\theta))$ are both strictly positive on Θ it is sufficient to show that

$$g^*(\theta) \geq h(\theta; \theta') := \frac{2v^*(\theta') + (\theta - \theta' - \dot{v}^*(\theta))\dot{v}^*(\theta')}{g^*(\theta')}, \quad (\theta, \theta') \in \Theta^2,$$

but, because we have that $h(\theta; \theta) = g(\theta)$ for all types, it is actually enough to show that the map $\theta' \mapsto h(\theta; \theta')$ is increasing on $[0, \theta]$, and decreasing on $[\theta, \theta_H]$. A direct

calculation shows that

$$[g^*(\theta')]^3 \frac{\partial h}{\partial \theta'}(\theta; \theta') = (\theta - \dot{v}^*(\theta) - \theta' + \dot{v}^*(\theta')) \ell(\theta'),$$

and (58) implies that the sign of this derivative is the same as that of

$$(\theta - \dot{v}^*(\theta)) - (\theta' - \dot{v}^*(\theta')).$$

Due to (55) we have that the function $\theta - \dot{v}^*(\theta)$ is non decreasing. Therefore, the sign of the above expression is the same as that of the difference $\theta - \theta'$. ■

Lemma B.8 *The map $\theta \mapsto F(\theta, v^*(\theta), \dot{v}^*(\theta))$ is non decreasing.*

Proof. A direct calculation shows that

$$\frac{dF}{d\theta}(\theta, v^*(\theta), \dot{v}^*(\theta)) = \left(\theta - \frac{\dot{v}^*(\theta)}{g^*(\theta)} \xi \right) \dot{v}^*(\theta) + \left(\frac{\xi}{g^*(\theta)} - 1 \right) \dot{v}^*(\theta).$$

Since the function $v^*(\theta)$ is increasing and convex by Lemma B.5 we only need to show that the bracketed terms are nonnegative. Consider the first term. Since

$$\theta - \frac{\dot{v}^*(\theta)}{g^*(\theta)} \xi = \Lambda(\theta; v^*(0))$$

we know from the proof of Lemma B.7 that this term is concave in θ , equal to zero on the boundary of the type space, and such that

$$\frac{d}{d\theta} \left(\theta - \frac{\dot{v}^*(\theta)}{g^*(\theta)} \xi \right) \Big|_{\theta=0} > 0.$$

This implies that this term is nonnegative throughout Θ . On the other hand, a direct calculation using the definition of the function $g^*(\theta)$ shows that

$$\frac{d}{d\theta} \left(\frac{\xi}{g^*(\theta)} - 1 \right) = -\frac{\dot{g}^*(\theta) \xi}{[g^*(\theta)]^2} = -\frac{\dot{v}^*(\theta)(1 - \dot{v}^*(\theta)) \xi}{[g^*(\theta)]^{5/2}}.$$

Since $v^*(0) \in \mathcal{C}_1$ we know from Lemma B.5 that $\dot{v}^*(\theta) \leq 1$ and $\dot{v}^*(\theta) \geq 0$. Therefore, the above expression is negative throughout the type space and the desired result now follows by observing that

$$\frac{g^*(\theta_H)}{\xi} = \frac{\dot{v}^*(\theta_H)}{\theta_H} = \int_0^{\theta_H} \dot{v}^*(\theta) \frac{d\theta}{\theta_H} \geq 1,$$

due to (12), the fundamental theorem of calculus and Lemma B.5. ■

Proof of Proposition 6. A direct calculation using the fact that $v^*(\theta)$ solves (10) subject to (11)–(12) shows that we have

$$\begin{aligned}\Delta(0) &= \Delta(\theta_H) = 0, \\ \dot{\Delta}(\theta) &= \frac{1}{g^*(\theta)\bar{\zeta}} \left(\frac{\bar{\zeta}}{2} - g^*(\theta) \right),\end{aligned}$$

and therefore $\dot{\Delta}(0) > 0$ since $v^*(0) < \frac{1}{8}\bar{\zeta}^2$ by Lemma B.6. In view Lemma B.5 we have that $g^*(\theta)$ is nonnegative and increasing. Therefore, the derivative $\dot{\Delta}(\theta)$ only changes sign once, and the desired result now follows from the fact that the function is equal to zero on the boundary of the type space. ■

Proof of Proposition 7. To establish the existence of a constant θ_2 with the required property we need to show that the function

$$\ell(\theta) := a(\pi(\theta, \phi^*(\theta))\phi_1^*(\theta) - q^\circ(\theta)\phi_1^\circ(\theta)) = g^*(\theta) - \frac{\bar{\zeta}}{2}$$

is first negative then positive. Since $\dot{v}^*(\theta) \leq 1$ and $\dot{v}^*(\theta) \geq 0$, by Lemma B.5 we have that the function $\ell(\theta)$ is increasing, and it is thus sufficient to show that it crosses the horizontal axis. Consider the function

$$\Delta(\theta) := \frac{\pi(\theta, \phi^*(\theta))\phi_2^*(\theta)}{\pi(\theta, \phi^*(\theta))\phi_1^*(\theta)} - \frac{q^\circ(\theta)\phi_2^\circ(\theta)}{q^\circ(\theta)\phi_1^\circ(\theta)} = \frac{\dot{v}^*(\theta)}{g^*(\theta)} - \frac{\theta}{\bar{\zeta}}.$$

As shown in the proof of Proposition 6, we have that the derivative of this function changes sign only once and the desired result now follows by observing that

$$\dot{\Delta}(\theta) = \frac{\ell(\theta)}{g^*(\theta)\bar{\zeta}}.$$

To show the existence of a constant θ_1 with the required property, consider the function

$$k(\theta) := a(\pi(\theta, \phi^*(\theta))\phi_2^*(\theta) - q^\circ(\theta)\phi_2^\circ(\theta)) = \dot{v}^*(\theta) - \frac{\theta}{2}$$

Combining Lemmas B.5 and B.6, we deduce that this function is increasing, convex, and such that

$$\dot{k}(0) = \dot{v}^*(0) - \frac{1}{2} < 0 = k(0).$$

Therefore, the function $k(\theta)$ crosses the horizontal axis at most once and the existence of a constant θ_1 with the required property now follows by observing that, due to (12),

the increase of the function of $g^*(\theta)$ and the definition of θ_2 , we have

$$k(\theta_H) = v^*(\theta_H) - \frac{\theta_H}{2} = \frac{\theta_H}{\xi} \left(g^*(\theta_H) - \frac{\xi}{2} \right) \geq \frac{\theta_H}{\xi} \left(g^*(\theta_2) - \frac{\xi}{2} \right) = 0$$

Let us now show that the constants θ_1 and θ_2 are such that $\theta_1 \leq \theta_2$. Since

$$\frac{\pi(\theta, \phi^*(\theta))\phi_2^*(\theta)}{\pi(\theta, \phi^*(\theta))\phi_1^*(\theta)} \leq \frac{q^\circ(\theta)\phi_2^\circ(\theta)}{q^\circ(\theta)\phi_1^\circ(\theta)}$$

for all $\theta \in \Theta$, by Proposition 6 it follows from the definition of θ_1 that we have

$$\begin{aligned} \pi(\theta_1, \phi^*(\theta_1))\phi_2^*(\theta_1) &\leq \pi(\theta_1, \phi^*(\theta_1))\phi_1^*(\theta_1) \frac{q_2(\theta_1)\phi_2^\circ(\theta_1)}{q_2(\theta_1)\phi_1^\circ(\theta_1)} \\ &= \pi(\theta_1, \phi^*(\theta_1))\phi_1^*(\theta_1). \end{aligned}$$

Therefore, $k(\theta_1) \leq 0$ and the desired conclusion now follows from the first part of the proof. The remaining claims in the statement follow from Lemma B.9 below. \blacksquare

Lemma B.9 *There exists $\bar{\theta} \in [\theta_1, \theta_2]$ such that $v^*(\theta) \leq v^\circ(\theta)$ if and only if $\theta \leq \bar{\theta}$.*

Proof. Consider the function defined by

$$m(\theta) := v^*(\theta) - v^\circ(\theta) = v^*(\theta) - \frac{1}{8}\|\xi(\theta)\|^2.$$

Since $v^*(0) < \xi^2/8$ by Lemma B.6, we have that $m(0) < 0$. On the other hand, the result of Proposition 7 implies that we have

$$\begin{aligned} m(\theta_H) &= v^*(\theta_H) - \frac{1}{8}\|\xi(\theta_H)\|^2 \\ &= \frac{1}{2}[g^*(\theta_H)]^2 + \frac{1}{2}[v^*(\theta_H)]^2 - \frac{1}{8}\|\xi(\theta_H)\|^2 \\ &\geq \frac{1}{2}[\xi/2]^2 + \frac{1}{2}[\theta_H/2]^2 - \frac{1}{8}\|\xi(\theta_H)\|^2 = 0, \end{aligned}$$

and it follows from the intermediate value theorem that there exist $\bar{\theta} \in \Theta$ such that $m(\bar{\theta}) = 0$. To complete the proof it is now sufficient to show that this point is unique and lies between θ_1 and θ_2 . A direct calculation gives

$$0 = \dot{m}(\bar{\theta}) - \left(\dot{v}(\bar{\theta}) - \frac{1}{4}\bar{\theta} \right) = \ddot{m}(\bar{\theta}) - \left(\ddot{v}(\bar{\theta}) - \frac{1}{4} \right),$$

and, as shown in the proof of Lemma B.5, we have that $\ddot{v}^*(\theta)$ is non decreasing. It follows that two cases may occur. If $\ddot{m}(\bar{\theta}) > 0$, then the function $m(\theta)$ is convex and

therefore increasing, which implies that it can cross the horizontal axis at most once. On the contrary, if $\dot{m}(0) \leq 0$ then the derivative $\dot{m}(\theta)$ changes sign at most once and the existence of unique crossing point follows. Finally, using the result of Proposition 7 we obtain that

$$\begin{aligned} v^*(\theta_1) &= \frac{1}{2}[\dot{v}^*(\theta_1)]^2 + \frac{1}{2}[g^*(\theta_1)]^2 = \frac{1}{2}[\dot{v}^*(\theta_1)]^2 + \frac{1}{8}\xi^2 \\ &\leq \frac{1}{2}[\theta_1/2]^2 + \frac{1}{8}\xi^2 = \frac{1}{8}\|\xi(\theta_1)\|^2, \\ v^*(\theta_2) &= \frac{1}{2}[\dot{v}^*(\theta_2)]^2 + \frac{1}{2}[g^*(\theta_2)]^2 = \frac{1}{2}[\theta_2/2]^2 + \frac{1}{2}[g^*(\theta_2)]^2 \\ &\geq \frac{1}{2}[\theta_2/2]^2 + \frac{1}{2}[\xi/2]^2 = \frac{1}{8}\|\xi(\theta_2)\|^2, \end{aligned}$$

and the desired result now follows from the first part of the proof. \blacksquare

Lemma B.10 *The functions $\phi_1^*(\theta)$ and $\phi_2^*(\theta)$ are respectively decreasing and increasing with respect to the investor type.*

Proof. A direct calculation shows that we have

$$\dot{\phi}_1^*(\theta) = \frac{(\theta - \dot{v}^*(\theta))g^*(\theta)}{\xi F(\theta, v^*(\theta), \dot{v}^*(\theta))^2} \left([\dot{v}^*(\theta)]^2 - 2v^*(\theta)\ddot{v}^*(\theta) \right).$$

Since $\ddot{v}^*(\theta) \leq 1$, by Lemma B.5 and $\dot{v}^*(0) = 0$ we have that $\theta - \dot{v}^*(\theta) \geq 0$. Therefore, the sign of the above derivative depends on the sign of the bracketed term on the right hand side and we know from the proof of Theorem 1 that this term is negative throughout the type space. Similarly, a direct calculation gives

$$\dot{\phi}_2^*(\theta) = \frac{(g^*(\theta) - \xi)}{g^*(\theta)F(\theta, v^*(\theta), \dot{v}^*(\theta))^2} \left([\dot{v}^*(\theta)]^2 - 2v^*(\theta)\ddot{v}^*(\theta) \right)$$

and the same argument as in the first part of the proof show that the sign of this quantity depends on that of the function $\xi - g^*(\theta)$. Because $\ddot{v}^*(\theta) \leq 1$, we have that this function is increasing and the desired result now follows by observing that

$$g^*(\theta_H) = (\xi/\theta_H)\dot{v}^*(\theta_H) \leq \xi$$

as a result of (12) and the fact that $\dot{v}^*(\theta) \leq \theta$ for all $\theta \in \Theta$. \blacksquare

Proof of Proposition 4. This directly follows by combining Lemmas B.8 and B.10. \blacksquare

Proof of Proposition 5. By Theorem 1 we have that

$$M = \sup_{v \in \Phi} \int_0^{\theta_H} F(\theta, v(\theta), \dot{v}(\theta)) \frac{d\theta}{\theta_H} = \int_0^{\theta_H} F(\theta, v^*(\theta), \dot{v}^*(\theta)) \frac{d\theta}{\theta_H}$$

Therefore, the envelope theorem implies that

$$\frac{dM}{d\zeta} = \int_0^{\theta_H} \frac{dF}{d\zeta}(\theta, v^*(\theta), \dot{v}^*(\theta)) \frac{d\theta}{\theta_H} = \int_0^{\theta_H} g^*(\theta) \frac{d\theta}{\theta_H},$$

and the required monotonicity in ζ follows from the nonnegativity of $g^*(\theta)$. Similarly, an application of the envelope theorem shows that

$$\frac{dM}{d\theta_H} = \frac{1}{\theta_H^2} \int_0^{\theta_H} \left(F(\theta_H, v^*(\theta_H), \dot{v}^*(\theta_H)) - F(\theta, v^*(\theta), \dot{v}^*(\theta)) \right) \frac{d\theta}{\theta_H}$$

and the desired monotonicity follows from Lemma B.8. ■

Proof of Lemma 2. By Proposition 1 we have that there exists an incentive compatible fund loading function ϕ_U that implements the same indirect utilities as the unbundling solution. Therefore $M_0 \leq M$ and the desired result will follow once we show that the inequality is strict. Let

$$v_U(\theta) = \frac{1}{2\|\phi_U(\theta)\|^2} \left(\phi_U(\theta)^\top \zeta(\theta) - 1 \right)_+^2 = \frac{1}{8}\zeta^2 + \frac{1}{2} \left(\theta - \frac{\theta_H}{3} \right)_+^2$$

denote the indirect utility that investors derive when offered ϕ_U . By Lemma 1 we have that $v_U \in \mathcal{V}$ and it thus follows from Theorem 1 that

$$M_0 = \int_{\Theta} F(\theta, v_U(\theta), \dot{v}_U(\theta)) \frac{d\theta}{\theta_H} \leq M = \sup_{v \in \mathcal{V}} \int_{\Theta} F(\theta, v(\theta), \dot{v}(\theta)) \frac{d\theta}{\theta_H}.$$

Since the supremum on the right is uniquely attained by the function v^* , it now suffices to show that the functions v^* and v_U differ over an open subset of the type space and this property follows by observing that over the interval $(0, \theta_H/3)$ the function v_U is constant whereas the function v^* is strictly increasing. ■

C Proof of the results in Section 3

C.1 Incentive compatible menus

Proof of Propositions 8 and 9. The proofs of these propositions are similar to those of Propositions 1 and 2. We omit the details. ■

Proof of Lemma 4. The proof is similar to that of Lemma 1. We omit the details. ■

C.2 The relaxed problem

Proof of Lemma 3. Since $\Phi_1 \subseteq \Phi$, we have that $M_1 \leq M$ and it follows that it is sufficient to show that v^* satisfies (19). The definition of the function v^* and the results of Lemmas B.5 and B.6 imply that $\ddot{v}^*(\theta) \leq 1$ for all $\theta \in \Theta$. Therefore, the function $b^*(\theta) := 2v^*(\theta) - [\dot{v}^*(\theta)]^2$ is non decreasing, and the desired result follows by noting that under the assumption of the statement we have $b^*(0) \geq (\xi - \gamma_1)^2$. ■

Proof of Lemma 5. Assume that the conditions of the statement hold true and pick an arbitrary $w \in \mathcal{V}_1$. Combining Lemma C.1 below with (22), the definition of \mathcal{V}_1 and the nonnegativity of the Lagrange multiplier we deduce that

$$\begin{aligned} \delta(v; w) &:= \int_0^{\theta_H} \left(F(\theta, w(\theta), \dot{w}(\theta)) - F(\theta, v(\theta), \dot{v}(\theta)) \right) d\theta \\ &\leq \int_0^{\theta_H} \left(H^\lambda(\theta, w(\theta), \dot{w}(\theta)) - H^\lambda(\theta, v(\theta), \dot{v}(\theta)) \right) d\theta \\ &\leq \int_0^{\theta_H} \left((w(\theta) - v(\theta)) H_{v(\theta)}^\lambda(\theta) + (\dot{w}(\theta) - \dot{v}(\theta)) H_{\dot{v}(\theta)}^\lambda(\theta) \right) d\theta \\ &= \sum_{n=0}^N \int_{\theta_n}^{\theta_{n+1}} \left((w(\theta) - v(\theta)) H_{v(\theta)}^\lambda(\theta) + (\dot{w}(\theta) - \dot{v}(\theta)) H_{\dot{v}(\theta)}^\lambda(\theta) \right) d\theta, \end{aligned} \quad (59)$$

where $\theta_0 = 0$, $\theta_{N+1} = \theta_H$, the N -tuple $(\theta_n)_{n=1}^N \in \text{int}(\Theta)^N$ identifies the points of discontinuity of the multiplier, and we have set

$$H_k^\lambda(\theta) = H_k^\lambda(\theta, v(\theta), \dot{v}(\theta)), \quad k \in \{v(\theta), \dot{v}(\theta)\}.$$

By assumption we have that the functions v , \dot{v} , w , and λ are absolutely continuous on the interval (x_n, x_{n+1}) . Therefore, the functions $\dot{w} - \dot{v}$ and

$$H_{\dot{v}(\theta)}^\lambda(\theta) = F_{\dot{v}(\theta)}(\theta, v(\theta), \dot{v}(\theta)) - 2\dot{v}(\theta)\lambda(\theta)$$

are, respectively, Lebesgue integrable and absolutely continuous on Θ ; and we can thus integrate by parts on the right hand side of (59) to obtain that

$$\begin{aligned}
\delta(v; w) &\leq \sum_{n=0}^N \left\{ (w(\theta) - v(\theta)) H_{\dot{v}(\theta)}^\lambda(\theta) \Big|_{\theta_n}^{\theta_{n+1}} \right. \\
&\quad \left. + \int_{\theta_n}^{\theta_{n+1}} (w(\theta) - v(\theta)) \left[H_{v(\theta)}^\lambda(\theta) - \frac{d}{d\theta} H_{\dot{v}(\theta)}^\lambda(\theta) \right] d\theta \right\} \\
&= \sum_{n=0}^N (w(\theta) - v(\theta)) H_{\dot{v}(\theta)}^\lambda(\theta) \Big|_{\theta_n}^{\theta_{n+1}} \\
&= (w(\theta) - v(\theta)) H_{\dot{v}(\theta)}^\lambda(\theta) \Big|_0^{\theta_H} - \sum_{n=1}^N (w(\theta_n) - v(\theta_n)) \Delta H_{\dot{v}(\theta)}^\lambda(\theta_n) \\
&= (w(\theta) - v(\theta)) H_{\dot{v}(\theta)}^\lambda(\theta) \Big|_0^{\theta_H} = 0,
\end{aligned}$$

where the first equality follows from (20), the second follows by expanding and then collapsing the sum, and the last follows from (21). \blacksquare

Remark C.1 Note that for the function

$$H_{\dot{v}(\theta)}^\lambda(\theta) = 2\dot{v}(\theta)\lambda(\theta) + F_{\dot{v}(\theta)}(\theta)$$

to be continuous it is necessary and sufficient that the optimal indirect utility be such that $\dot{v}(\theta) = 0$ at every point of discontinuity of the multiplier.

Lemma C.1 *Let $(\theta, \lambda) \in \Theta \times AC_p^*(\Theta; \mathbf{R}_+)$ be fixed. Then the map $(v, \dot{v}) \mapsto H^\lambda(\theta, v, \dot{v})$ is strictly concave on the set of pairs such that $c(v, \dot{v}) \geq 0$.*

Proof. This follows by verifying that the determinant and trace of the Hessian matrix are respectively strictly positive and strictly negative. We omit the details. \blacksquare

Proof of Proposition 10 when $\gamma_1 \leq \xi/3$. To establish the result it suffices to show that one can construct a Lagrange multiplier $\lambda \in AC_p^*(\Theta; \mathbf{R}_+)$ such that the pair (λ, v_1^*) satisfies the conditions of Lemma 5. As is easily seen we have that the candidate optimizer belongs to $C_p^2(\Theta; \mathbf{R})$ and satisfies

$$c(v_1^*(\theta), \dot{v}_1^*(\theta)) = 0, \quad \theta \in \Theta.$$

Therefore, (22) holds. On the other hand, substituting the candidate optimizer into

(20) and (21) shows that the Lagrange multiplier must be chosen in such a way that

$$\begin{aligned} 0 &= \frac{\gamma_1}{\xi - \gamma_1} + 2\lambda(\theta) - 2, & \theta &\leq \frac{\theta_H}{3} \\ 0 &= \frac{\gamma_1}{\xi - \gamma_1} + 2\lambda(\theta) - \frac{1}{2} + \lambda(\theta) \left(\theta - \frac{\theta_H}{3} \right), & \theta &> \frac{\theta_H}{3}, \end{aligned}$$

and

$$0 = \frac{\gamma_1}{\xi - \gamma_1} + 2\lambda(\theta_H) - \frac{1}{2}.$$

A direct calculation shows that the unique solution to these equations is piecewise constant and explicitly given by

$$\lambda(\theta) := 1 - \frac{\gamma_1}{2(\xi - \gamma_1)} - \mathbf{1}_{\{3\theta > \theta_H\}} \frac{3}{4}.$$

Since $\gamma_1 \leq \xi/3$, we have that $\lambda(\theta)$ is nonnegative for all $\theta \in \Theta$. Therefore, it now only remains to establish that the function

$$\theta \mapsto H_{\dot{v}_1^*(\theta)}^\lambda(\theta, v_1^*(\theta), \dot{v}_1^*(\theta)) = \theta - \dot{v}_1^*(\theta) \left(2\lambda(\theta) + \frac{\xi}{\sqrt{2v_1^*(\theta) - [\dot{v}_1^*(\theta)]^2}} \right)$$

is continuous on Θ but this property follows from Remark C.1, the smoothness of the candidate optimizer, and the fact that $\dot{v}_1^*(\theta_H/3) = 0$. \blacksquare

Lemma C.2 *For every $\gamma_1 \leq \gamma_1^*$ there exists a unique solution $(w, \theta^*) \in C_p^2(\Theta; \mathbf{R}) \times \Theta$ to the free boundary problem defined by (24), (25), and (26).*

Proof. We start by observing that (w, θ^*) is a solution to the free boundary problem if and only if $m(x) := w(x + \theta^*)$ solves the initial value problem

$$m(x) (1 + \dot{m}(x)) - [\dot{m}(x)]^2 = \frac{3}{2\xi} \left(2m(x) - [\dot{m}(x)]^2 \right)^{\frac{3}{2}}, \quad (60)$$

$$\dot{m}(0) = m(0) - \frac{1}{2}(\xi - \gamma_1)^2 = 0, \quad (61)$$

and the constant $\theta^* \in \Theta$ solves

$$Q(\theta^*) := \theta_H - \frac{\xi \dot{m}(\theta_H - \theta^*)}{\sqrt{2m(\theta_H - \theta^*) - [\dot{m}(\theta_H - \theta^*)]^2}}. \quad (62)$$

Since $\gamma_1 \leq \gamma_1^*$, we have that

$$\frac{1}{2}(\xi - \gamma_1)^2 \geq \frac{1}{2}(\xi - \gamma_1^*)^2 = v^*(0) > \frac{1}{18}\xi^2.$$

Therefore, it follows from Lemma B.5 that the unique classical solution to (60) subject to (61) is given by $m(x) = v(x; q_1)$ with $q_1 = \frac{1}{2}(\xi - \gamma_1)^2$ and it now only remains to show that (62) admits a unique solution.

Since the function Q is continuous on Θ and $Q(\theta_H) = \theta_H > 0$, the existence claim will follow from the intermediate value theorem once we show that $Q(0) \leq 0$. To this end, consider the continuously differentiable function

$$S(\theta) := \theta - \frac{\xi \dot{v}(\theta; q_1)}{\sqrt{2v(\theta, q_1) - [\dot{v}(\theta; q_1)]^2}}$$

and observe that, since $Q(0) = S(\theta_H)$, it is sufficient to prove that $S(\theta) \leq 0$ in a left neighbourhood of θ_H . Differentiating the above expression and using the fact that the function $v(\theta; q_1)$ solves (60) shows that

$$\dot{S}(\theta) = \frac{\xi}{\sqrt{2v(\theta, q_1) - [\dot{v}(\theta; q_1)]^2}} - 2. \quad (63)$$

To proceed further we distinguish three cases. If the fee rate on the outside fund is such that $q_1 \in \mathcal{C}_2$, then we know from the proof of Lemma B.5 that

$$\inf_{\theta \in \Theta} \sqrt{2v(\theta, q_1) - [\dot{v}(\theta; q_1)]^2} \geq \frac{2}{3}\xi.$$

This implies that we have $\dot{S}(\theta) \leq -\frac{1}{2}$ for all $\theta \in \Theta$ and the desired result follows by noting that $S(0) = 0$. Assume next that the fee rate is such that $q_1 \in [\frac{1}{8}, \frac{2}{9}]\xi^2$. In this case we know from Lemma B.5 that $\ddot{v}(\theta; q_1) \leq 1$ for all $\theta \in \Theta$. This implies that the derivative in (63) is decreasing and the desired result follows by noting that

$$q_1 \geq \frac{1}{8}\xi^2 \implies \dot{S}(\theta) \leq \dot{S}(0) = \frac{\xi}{\sqrt{2q_1}} - 2 \leq 0 = S(0).$$

Finally, assume that the fee rate on the familiar asset is such that $q_1 < \frac{1}{8}\xi^2$, and denote by $q(\theta)$ the function that describes the unique strictly positive solution to (53) in \mathcal{C}_1 . As shown in the proof of Lemma B.6, this function is continuous, non-increasing, and equal to $\frac{1}{8}\xi^2$ at the origin. By continuity this implies that we have $q_1 = q(\hat{\theta})$ for some

strictly positive type $\hat{\theta}$ such that

$$S(\hat{\theta}) = \hat{\theta} - \frac{\xi \dot{v}(\hat{\theta}; q(\hat{\theta}))}{\sqrt{2v(\hat{\theta}, q(\hat{\theta})) - [\dot{v}(\hat{\theta}; q(\hat{\theta}))]^2}} = 0.$$

On the other hand, since $q_1 < \frac{1}{8}\xi^2 < \frac{2}{9}\xi^2$ we know from Lemma B.5 that $\dot{v}(\theta; q_1) \leq 1$ for all $\theta \in \Theta$ and it follows that the derivative in (63) is decreasing. Using this property in conjunction with the fact that

$$q_1 < \frac{1}{8}\xi^2 \implies \dot{S}(0) = -2 + \frac{\xi}{\sqrt{2q_1}} > 0 = S(\hat{\theta}),$$

we then deduce that $\dot{S}(\hat{\theta}) < 0$, and it follows that $S(\theta) \leq S(\hat{\theta}) = 0$ for all $\theta \geq \hat{\theta}$. To complete the proof it now remains to establish uniqueness. A direct calculation using (62) and the fact that the function $v(\theta; q_1)$ solves (60) implies that

$$Q'(\theta) = 3 - \frac{\xi}{\sqrt{2v(\theta, q_1) - [\dot{v}(\theta; q_1)]^2}} = 1 - \dot{S}(\theta).$$

If $q_1 \geq \frac{1}{8}\xi^2$ then the first part of the proof shows that we have $\dot{S}(\theta) \leq 0$ for all $\theta \in \Theta$. This implies that the function $Q(\theta)$ is strictly increasing and the required uniqueness follows. On the other hand, if we have $v^*(0) < q_1 < \frac{1}{8}\xi^2$, then we know from the first part of the proof that $Q'(\theta)$ is decreasing and has initial value

$$Q'(0) = 3 - \frac{\xi}{\sqrt{2q_1}} \geq 3 - \frac{\xi}{\sqrt{2\xi^2/18}} = 0.$$

Two cases may then occur: either $Q'(\theta_H) \geq 0$, in which case the function $Q(\theta)$ is strictly increasing, or $Q'(\theta_H) < 0$, in which case the function $Q(\theta)$ is inverse U -shaped with a maximum whose value exceeds $Q(\theta_H)$. The required uniqueness follows by observing that in both cases the function crosses the horizontal axis only once. ■

Proof of Proposition 10 when $\gamma_1 \in (\xi/3, \gamma_1^*]$. To establish the result it suffices to show that one can construct a Lagrange multiplier $\lambda \in AC_p^*(\Theta; \mathbf{R}_+)$ such that the pair (λ, v_1^*) satisfies the conditions of Lemma 5. As is easily seen we have that the candidate optimizer belongs to $C_p^2(\Theta; \mathbf{R})$ and satisfies

$$\begin{aligned} c(v_1^*(\theta), \dot{v}_1^*(\theta)) &= \mathbf{1}_{\{\theta > \theta^*\}} \left(2w(\theta) - [\dot{w}(\theta)]^2 - 2q_1 \right) \\ &= \mathbf{1}_{\{\theta > \theta^*\}} \left(2v(\theta - \theta^*; q_1) - [\dot{v}(\theta - \theta^*; q_1)]^2 - 2q_1 \right), \end{aligned}$$

where the second equality follows from the construction in the proof of Lemma C.2. Since $\gamma_1 \geq \frac{1}{3}\xi$, we have $q_1 \leq \frac{2}{9}\xi^2$. Therefore, the result of Lemma B.5 implies that the right hand side of the previous expression is strictly increasing in $\theta \in [\theta^*, \theta_H]$ and it follows that we have

$$c(v_1^*(\theta), \dot{v}_1^*(\theta)) > c(v_1^*(\theta^*), \dot{v}_1^*(\theta^*)) = 2(v(0; q_1) - q_1) = 0, \quad \theta > \theta^*,$$

which establishes that the candidate optimizer lies in \mathcal{V}_1 and shows that a necessary condition for (22) is that $\lambda(\theta) = 0$ for all $\theta > \theta^*$. On the other hand, using the fact that the function $v(\theta; q)$ solves (24) and substituting into (20) and (21) shows that the Lagrange multiplier must be chosen in such a way that

$$0 = \frac{\gamma_1}{\xi - \gamma_1} + 2\lambda(\theta) - 2, \quad \theta \leq \theta^*.$$

Solving that equation shows the Lagrange multiplier is given by

$$\lambda(\theta) := \mathbf{1}_{\{\theta \leq \theta^*\}} \left(1 - \frac{\gamma_1}{2(\xi - \gamma_1)} \right).$$

Since $\gamma_1 \leq \xi/3$, we have that $\lambda(\theta)$ is nonnegative for all $\theta \in \Theta$. Therefore, it now only remains to establish that the function

$$\theta \mapsto H_{\dot{v}_1^*(\theta)}^\lambda(\theta, v_1^*(\theta), \dot{v}_1^*(\theta)) = \theta - \dot{v}_1^*(\theta) \left(2\lambda(\theta) + \frac{\xi}{\sqrt{2v_1^*(\theta) - [\dot{v}_1^*(\theta)]^2}} \right)$$

is continuous on Θ but this property follows from Remark C.1, the smoothness of the candidate optimizer, and the fact that $\dot{v}_1^*(\theta^*) = \dot{v}(0; q_1) = 0$. ■

C.3 The optimal fund menu

Proof of Theorem 2. Since $v_1^* \in C_p^2(\Theta; \mathbf{R})$, we have $\phi_1^* \in AC(\Theta; \mathbf{R}^2)$. Therefore, Lemma 4 and Proposition 10 imply that

$$M_1 = \sup_{\phi \in \Phi} I(\phi) \leq \frac{V_1}{\theta_H} = \int_{\Theta} F(\theta, \dot{v}_1^*(\theta), \dot{v}_1^*(\theta)) \frac{d\theta}{\theta_H} = I(\phi_1^*)$$

and the statement will follow once we show that the fund loading function ϕ_1^* belongs to Φ_1 and satisfies (28). We distinguish two cases depending on the fee rate of the outside fund. Assume first that we have $\gamma_1 \leq \frac{1}{3}\xi$ so that the function $v_1^*(\theta)$ is given by

(23). In this case the result follows by observing that we have

$$F(\theta, v_1^*(\theta), \dot{v}_1^*(\theta)) = \frac{\gamma_1}{a} (\zeta - \gamma_1) + \frac{\theta^*}{a} (\theta - \theta^*)_+ > 0,$$

as well as

$$\zeta - \gamma_1 = \frac{\phi_{11}^*(\theta)}{\|\phi_1^*(\theta)\|^2} \left(\phi_1^*(\theta)^\top \zeta(\theta) - 1 \right)_+,$$

and

$$\begin{aligned} \phi_1^*(\theta')^\top \zeta(\theta) - 1 - \frac{\phi_1^*(\theta)^\top \phi_1^*(\theta')}{\|\phi_1^*(\theta)\|^2} \left(\phi_1^*(\theta)^\top \zeta(\theta) - 1 \right)_+ \\ = \mathbf{1}_{\{\theta \leq \theta^* < \theta'\}} \frac{(\theta^* - \theta)(\theta^* - \theta')}{\gamma_1(\zeta - \gamma_1) + \theta^*(\theta' - \theta^*)} \leq 0 \end{aligned}$$

for all $(\theta, \theta') \in \Theta^2$. Assume next that the fee rate $\gamma_1 \in (\frac{1}{3}\zeta^2, \gamma_1^*]$, and let us start by establishing the validity of (28). Since $v_1^*(\theta)$ is piecewise smooth, we have that the mapping defined by

$$\theta \mapsto F_1^*(\theta) := F(\theta, v_1^*(\theta), \dot{v}_1^*(\theta))$$

is absolutely continuous. On the other hand, since $\gamma_1 > \frac{1}{3}\zeta$ it follows from (27) and the proofs of Lemmas B.5 and C.2 that we have

$$\min \left\{ \dot{v}_1^*(\theta), 1 - \ddot{v}_1^*(\theta), \dot{v}_1^*(\theta), \ddot{v}_1^*(\theta) \right\} \geq 0, \quad \theta \in [\theta^*, \theta_H]. \quad (64)$$

Using this property in conjunction with the fundamental theorem of calculus, we then deduce that

$$\begin{aligned} \dot{F}_1^*(\theta) &= \mathbf{1}_{\{\theta \geq \theta^*\}} \left(\theta \ddot{v}_1^*(\theta) - \dot{v}_1^*(\theta) + \dot{v}_1^*(\theta)(1 - \ddot{v}_1^*(\theta)) \frac{\zeta}{g_1^*(\theta)} \right) \\ &\geq \mathbf{1}_{\{\theta \geq \theta^*\}} (\theta \ddot{v}_1^*(\theta) - \dot{v}_1^*(\theta)) = \int_{\theta^*}^{\max\{\theta, \theta^*\}} (\ddot{v}_1^*(\theta) - \ddot{v}_1^*(x)) dx \geq 0. \end{aligned}$$

for the function

$$g_1^*(\theta) := \sqrt{2v_1^*(\theta) - [\dot{v}_1^*(\theta)]^2}.$$

This shows that the absolutely continuous function $F_1^*(\theta)$ is nondecreasing throughout the type space and (28) now follows by observing that

$$F(\theta, v_1^*(\theta), \dot{v}_1^*(\theta)) = \gamma_1(\xi - \gamma_1) > 0, \quad \theta \in [0, \theta^*].$$

To complete the proof we now need to show that the fund loading function $\phi_1^*(\theta)$ is incentive compatible. A direct calculation using (29) shows that

$$\xi - \gamma_1 - \frac{\phi_{11}^*(\theta)}{\|\phi_1^*(\theta)\|^2} \left(\phi_1^*(\theta)^\top \xi(\theta) - 1 \right)_+ = \mathbf{1}_{\{\theta \geq \theta^*\}} (g_1^*(\theta^*) - g_1^*(\theta)),$$

and (18) follows by noting that, as a result of (64), the function $g_1^*(\theta)$ is nondecreasing on the interval $[\theta^*, \theta_H]$. On the other hand, using (64) and proceeding as in the proof of Theorem 1 shows that the validity of (4) is equivalent to

$$g_1^*(\theta) \geq h(\theta, \theta') := \frac{2v_1^*(\theta') + (\theta - \theta' - \dot{v}_1^*(\theta)) \dot{v}_1^*(\theta')}{g_1^*(\theta')}, \quad (\theta, \theta') \in \Theta^2. \quad (65)$$

To prove this inequality we start by decomposing the set Θ^2 into the union of the disjoint subsets $(\Theta_i)_{i=1}^4$ defined by

$$\begin{aligned} \Theta_1 &:= \{(\theta, \theta') \in \Theta^2 : \max\{\theta, \theta'\} \leq \theta^*\}, \\ \Theta_2 &:= \{(\theta, \theta') \in \Theta^2 : \theta \leq \theta^* \text{ and } \theta' > \theta^*\}, \\ \Theta_3 &:= \{(\theta, \theta') \in \Theta^2 : \theta > \theta^* \text{ and } \theta' \leq \theta^*\}, \end{aligned}$$

and

$$\Theta_4 := \{(\theta, \theta') \in \Theta^2 : \min\{\theta, \theta'\} > \theta^*\}.$$

On the set Θ_1 the inequality holds since

$$g_1^*(\theta) = h(\theta, \theta') = \sqrt{2q_1} = \xi - \gamma_1, \quad (\theta, \theta') \in \Theta_1.$$

On the set Θ_3 the inequality boils down to

$$g_1^*(\theta) \geq g_1^*(\theta^*) = \sqrt{2q_1} = \xi - g_1, \quad \theta > \theta^*,$$

which is satisfied because $g_1^*(\theta)$ is non decreasing on $[\theta^*, \theta_H]$ as a result of (64). On the set Θ_2 we have that

$$g_1^*(\theta) - h(\theta, \theta') = \zeta - \gamma_1 - \frac{(\zeta - \gamma_1)^2 + (\theta - \theta')\dot{v}_1^*(\theta')}{g_1^*(\theta')}$$

is strictly decreasing with respect to θ and it follows that the validity of (65) on that set is equivalent to

$$h(\theta^*, \theta') \leq \zeta - \gamma_1, \quad \theta' > \theta^*. \quad (66)$$

Differentiating the right hand side of (65) gives

$$\frac{\partial h}{\partial \theta'}(\theta, \theta') = (\theta - \dot{v}_1^*(\theta) - \theta' + \dot{v}_1^*(\theta')) \left[\frac{2v_1^*(\theta')\ddot{v}_1^*(\theta') - [\dot{v}_1^*(\theta')]^2}{g_1^*(\theta')^3} \right].$$

Combining (64) with the fundamental theorem of calculus we deduce that

$$2v_1^*(\theta')\dot{v}_1^*(\theta') - [\dot{v}_1^*(\theta')]^2 = \int_{\theta^*}^{\theta'} 2v_1^*(x)\ddot{v}_1^*(x)dx \geq 0 \quad (67)$$

for all $\theta' \geq \theta^*$. On the other hand, since $\dot{v}_1^*(\theta^*) = 0$ it follows from (64) and the fundamental theorem of calculus that

$$\theta^* - \theta' + \dot{v}_1^*(\theta') = \theta^* - \theta' + \int_{\theta^*}^{\theta'} \ddot{v}_1^*(x)dx \leq 0$$

for all $\theta' \geq \theta$. This shows that $h(\theta^*, \theta')$ is decreasing in θ' and (66) now follows by observing that

$$h(\theta^*, \theta^*) = \frac{2v_1^*(\theta^*) - [\dot{v}_1^*(\theta^*)]^2}{g_1^*(\theta^*)} = g_1^*(\theta^*) = \zeta - \gamma_1.$$

Consider finally the set Θ_4 . Since $h(\theta, \theta) = g_1^*(\theta)$, it is sufficient to show that for any fixed $\theta > \theta^*$ the function $h(\theta, \theta')$ reaches a maximum over $(\theta^*, \theta_H]$ at the point $\theta' = \theta$. In view of (67) we have that the sign of $\frac{\partial h}{\partial \theta'}$ is determined by the sign of

$$(\theta - \dot{v}_1^*(\theta)) - (\theta' - \dot{v}_1^*(\theta')).$$

By (64) we have that $x - \dot{v}_1^*(x)$ is nondecreasing on $[\theta^*, \theta_H]$ and it follows that the above expression is nonnegative if and only if $\theta' \leq \theta$. ■

Proof of Proposition 11. Arguments similar to those of Sections 2.2 and 3.2 show that

under exclusivity the value function of the manager satisfies

$$\theta_H M_1 \leq \theta_H M_{1,E} \leq V_{1,E} \equiv \sup_{v \in \mathcal{V}_{1,E}} \int_{\Theta} F(\theta, v(\theta), \dot{v}(\theta)) d\theta, \quad (\mathcal{R}_{1,E})$$

where $\mathcal{V}_{1,E}$ denotes the set of functions $v \in AC(\Theta; \mathbf{R})$ that satisfy (30). Consider the candidate optimizer $v_{1,E}^*(\theta)$ defined in the statement. As is easily seen we have that this function is absolutely continuous. On the other hand, the construction of the function $w(\theta)$ implies that

$$\mathbf{1}_{\{\theta \geq \theta^*\}} (w(\theta) - v(\theta - \theta^*; q_1)) = 0 \quad \text{with} \quad q_1 \equiv \frac{1}{2}(\bar{\xi} - \gamma_1)^2.$$

Therefore, it follows from Lemma B.5 that $v_{1,E}^*(\theta)$ is nondecreasing on $[\theta^*, \theta_H]$ and, since $v_{1,E}^*(\theta) = q_1$ for all $\theta \leq \theta^*$, we conclude that $v_{1,E}^* \in \mathcal{V}_{1,E}$. To show that it attains the supremum consider the multiplier

$$\lambda_E(\theta) \equiv \mathbf{1}_{\{\theta \leq \theta^*\}} \left(3 - \frac{\bar{\xi}}{\bar{\xi} - \gamma_1} \right) \geq 0,$$

where the inequality follows from the fact that $\gamma_1 \leq \gamma_1^* \leq \frac{2}{3}\bar{\xi}$. A direct calculation shows that the pair $(v_{1,E}^*, \lambda_E)$ satisfies all the conditions of Lemma C.3 below and it thus follows that we have

$$V_{1,E} = \int_{\Theta} F(\theta, v_{1,E}^*(\theta), \dot{v}_{1,E}^*(\theta)) d\theta.$$

To complete the proof we distinguish two cases depending on the level of the fee rate on the familiar asset. Assume first that $\gamma_1 > \frac{1}{3}\bar{\xi}$. In this case

$$0 = |v_1^*(\theta) - v_{1,E}^*(\theta)| = \|\phi_1^*(\theta) - \phi_{1,E}^*(\theta)\|, \quad \theta \in \Theta,$$

and the desired result now follows from $(\mathcal{R}_{1,E})$ and Theorem 2. Assume next that the fee rate $\gamma \leq \frac{1}{3}\bar{\xi}$ and consider the fund loading function defined in the statement. To complete the proof we now show that this function belongs to $\Phi_{1,E}$ but not to Φ_1 . To establish the former we need to show that

$$\max \left\{ q_1, \sup_{\theta' \in \Theta} \frac{1}{2} \left(\frac{\phi_{1,E}^*(\theta')^\top \bar{\xi}(\theta) - 1}{\|\phi_{1,E}^*(\theta')\|} \right)_+^2 \right\} \leq v_{1,E}^*(\theta), \quad \theta \in \Theta.$$

but, since $v_{1,E}^*(\theta) \geq q_1$ for all $\theta \in \Theta$, we have that the validity of this inequality is

equivalent to the requirement that

$$B(\theta, \theta') \equiv 4v_{1,E}^*(\theta)v_{1,E}^*(\theta') - (2v_{1,E}^*(\theta') + (\theta - \theta')\dot{v}_{1,E}^*(\theta'))_+^2$$

be nonnegative for all $(\theta, \theta') \in \Theta \times (\theta^*, \theta_H]$. Differentiating this function with respect to its first argument gives

$$\frac{dB}{d\theta} = \mathbf{1}_{\{\theta > \theta^*\}} 4\dot{v}_{1,E}^*(\theta)v_{1,E}^*(\theta') - 2\dot{v}_{1,E}^*(\theta') (2v_{1,E}^*(\theta') + (\theta - \theta')\dot{v}_{1,E}^*(\theta'))_+$$

Since the fee rate $\gamma_1 < \frac{1}{3}\bar{\zeta}$, we have that $q_1 \in \mathcal{C}_2$. Therefore, it follows from Lemma B.5 that for all types $\theta' > \theta^*$ we have

$$\ddot{v}_{1,E}^*(\theta') = \ddot{v}(\theta' - \theta^*, q_1) \geq 1 \quad (68)$$

and

$$\dot{v}_{1,E}^*(\theta') = \dot{v}_{1,E}^*(\theta^*) + \int_{\theta^*}^{\theta'} \ddot{v}_{1,E}^*(\theta) d\theta \geq \theta' - \theta^*. \quad (69)$$

Combining these properties with (26) shows that for all $(\theta, \theta') \in [\theta^*, \theta_H]^2$ we have

$$\begin{aligned} 2v_{1,E}^*(\theta') + (\theta - \theta')\dot{v}_{1,E}^*(\theta') &\geq 2v_{1,E}^*(\theta') + (\theta^* - \theta')\dot{v}_{1,E}^*(\theta') \\ &\geq 2v_{1,E}^*(\theta') - [\dot{v}_{1,E}^*(\theta')]^2 \\ &\geq 2v_{1,E}^*(\theta_H) - [\dot{v}_{1,E}^*(\theta_H)]^2 = \left(\frac{\bar{\zeta}}{\theta_H}\right)^2 [\dot{v}_{1,E}^*(\theta_H)]^2 \geq 0, \end{aligned}$$

and therefore

$$\begin{aligned} (\theta - \theta') \frac{dB}{d\theta} &= 2(\theta - \theta')^2 \left(2v_{1,E}^*(\theta') \frac{\dot{v}_{1,E}^*(\theta) - \dot{v}_{1,E}^*(\theta')}{\theta - \theta'} - [\dot{v}_{1,E}^*(\theta')]^2 \right) \\ &= 2(\theta - \theta')^2 \left(2v_{1,E}^*(\theta') - [\dot{v}_{1,E}^*(\theta')]^2 \right) \geq 0. \end{aligned}$$

This shows that for any given $\theta' > \theta^*$ the function $\theta \mapsto B(\theta, \theta')$ reaches the minimum of zero over $[\theta^*, \theta_H]$, and we now have to consider types such that $\theta \leq \theta^* < \theta'$. For such types we have that $\frac{dB}{d\theta}$ is negative or zero, and the desired property now follows from the fact that, as shown above, we have $B(\theta^*, \theta') \geq 0$ for all $\theta' > \theta^*$.

To complete the proof it now remains to show that we have $\phi_{1,E}^* \notin \Phi_{1,E}$. Proceeding as in the proof of Theorem 2 we have that on $[\theta^*, \theta_H]^2$ the validity of the non exclusive

incentive compatibility condition is equivalent to

$$g_{1,E}^*(\theta) \geq h(\theta, \theta') \equiv \frac{2v_{1,E}^*(\theta') + (\theta - \theta' - \dot{v}_{1,E}^*(\theta))\dot{v}_{1,E}^*(\theta')}{g_{1,E}^*(\theta)} \quad (70)$$

with

$$g_{1,E}^*(\theta) \equiv \sqrt{2v_{1,E}^*(\theta) - [\dot{v}_{1,E}^*(\theta)]^2}.$$

Combining (68) and (69) we deduce that

$$2v_{1,E}^*(\theta)\ddot{v}_{1,E}^*(\theta) - [\dot{v}_{1,E}^*(\theta)]^2 \geq 2v_{1,E}^*(\theta) - [\dot{v}_{1,E}^*(\theta)]^2 \geq 0, \quad \theta \geq \theta^*.$$

Therefore, the sign of

$$\frac{dh}{d\theta'} = (\theta - \dot{v}_{1,E}^*(\theta) - \theta' + \dot{v}_{1,E}^*(\theta')) \left(\frac{2v_{1,E}^*(\theta')\ddot{v}_{1,E}^*(\theta') - [\dot{v}_{1,E}^*(\theta')]^2}{g_{1,E}^*(\theta')^3} \right)$$

is determined by the sign of the first bracket on the right. Because of (69) we have that the function $\theta - \dot{v}_{1,E}^*(\theta)$ is decreasing. This implies that

$$(\theta - \theta') \frac{dh}{d\theta'}(\theta, \theta') \leq 0, \quad \theta^* \leq \min\{\theta, \theta'\},$$

and using this inequality in conjunction with the fact that $h(\theta, \theta) = g_{1,E}^*(\theta)$ we deduce that $g_{1,E}^*(\theta) < h(\theta, \theta')$ for all $\theta \neq \theta'$ in $[\theta^*, \theta_H]^2$. This shows that the inequality in (70) fails and the desired result follows. \blacksquare

Consider the function defined by

$$G^\lambda(\theta, v, p) \equiv F(\theta, v, p) + \lambda(\theta)(v - q_1)$$

and observe that, as a result of Lemma B.4, this function is strictly concave in v and p . The following lemma is the counterpart of Lemma 5 for the case where the manager can force investors to commit to a single fund.

Lemma C.3 *Let $(v, \lambda) \in \mathcal{V}_{1,E} \times AC_p^*(\Theta; \mathbf{R}_+)$ be such that $\dot{v} \in AC(\Theta; \mathbf{R})$, and denote by \mathcal{C}*

the set of points where the function λ is continuous. If

$$\begin{aligned} \left(G_{v(\theta)}^\lambda - \frac{d}{d\theta} G_{\dot{v}(\theta)}^\lambda \right) (\theta, v(\theta), \dot{v}(\theta)) &= 0, & \theta \in \mathcal{C}, \\ G_{\dot{v}(\theta)}^\lambda (\theta, v(\theta), \dot{v}(\theta)) &= 0, & \theta \in \{0, \theta_H\}, \\ \lambda(\theta)(v(\theta) - q_1) &= 0, & \theta \in \Theta, \end{aligned}$$

and $G_{\dot{v}(\theta)}^\lambda (\theta, v(\theta), \dot{v}(\theta))$ is continuous, then v attains the supremum in $(\mathcal{R}_{1,E})$.

Proof. The proof is similar to that of Lemma 5. We omit the details. ■

References

- William Adams and Janet Yellen. Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics*, 90(3):475–498, 1976. [2](#), [6](#), [24](#)
- Anat Admati and Paul Pfleiderer. Does it all add up? benchmarks and the compensation of active portfolio managers. *Journal of Business*, 70(3):323–350, 1997. [5](#)
- Mark Armstrong. Multiproduct nonlinear pricing. *Econometrica*, 64(1):51–75, 1996. [1](#), [6](#)
- Suleyman Basak and Anna Pavlova. Asset prices and institutional investors. *American Economic Review*, 103(5):1728–58, 2013. [5](#)
- Suleyman Basak, Anna Pavlova, and Alexander Shapiro. Optimal asset allocation and risk shifting in money management. *Review of Financial Studies*, 20:1583–1621, 2007. [5](#)
- Jonathan Berk and Richard Green. Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295, 2004. [6](#)
- Jonathan Berk, Jules Van Binsbergen, and Binying Liu. Matching capital and labor. *Journal of Finance*, 72(6):2467–2504, 2017. [6](#)
- Jonathan B. Berk and Jules van Binsbergen. Mutual funds in equilibrium. *Annual Review of Financial Economics*, 9(1):147–167, 2017. [3](#)
- Sudipto Bhattacharya and Paul Pfleiderer. Delegated portfolio management. *Journal of Economic Theory*, 36(1):1 – 25, 1985. [5](#)
- Michèle Breton, Julien Hugonnier, and Tarek Masmoudi. Mutual fund competition in the presence of dynamic flows. *Automatica*, 46:1176–1185, 2010. [5](#)
- David Brown and Youchang Wu. Mutual fund flows and crossfund learning within families. *Journal of Finance*, 71(1):383–424, 2016. [5](#), [6](#)
- Bruce Carlin and Gustavo Manso. Obfuscation, learning, and the evolution of investor sophistication. *Review of Financial Studies*, 24:754–785, 2011. [5](#), [9](#)
- Jennifer Carpenter. Does option compensation increase managerial risk appetite? *Journal of Finance*, 55(5):2311–2331, 2000. [5](#)
- Martijn Cremers and Antti Petajisto. How active is your fund manager? a new measure that predicts performance. *Review of Financial Studies*, 22(9):3329–3365, 2009. [3](#)

- Martijn Cremers, Miguel Ferreira, Pedro Matos, and Laura Starks. Indexing and active fund management: International evidence. *Journal of Financial Economics*, 120(3):539 – 560, 2016. [3](#)
- Domenico Cuoco and Ron Kaniel. Equilibrium prices in the presence of delegated portfolio management. *Journal of Financial Economics*, 101(2):264 – 296, 2011. [5](#)
- Jakša Cvitanić and Hao Xing. Asset pricing under optimal contracts. *Journal of Economic Theory*, 173:142 – 180, 2018. ISSN 0022-0531. [5](#)
- Thomas Dangl, Youchang Wu, and Josef Zechner. Market discipline and internal governance in the mutual fund industry. *Review of Financial Studies*, 21(5):2307–2343, 2008. [5](#)
- Sanjiv Das and Rangarajan Sundaram. On the regulation of fee structures in mutual funds. Working paper, National Bureau of Economic Research, 1998. [2](#)
- Sanjiv Das and Rangarajan Sundaram. Fee speech: Signaling, risk-sharing, and the impact of fee structures on investor welfare. *Review of Financial Studies*, 15(5):1465–1497, 2002. [5](#)
- Philip Dybvig, Heber Farnsworth, and Jennifer Carpenter. Portfolio performance and agency. *Review of Financial Studies*, 23(1):1–23, 2010. [5](#)
- Martin Gruber. Another puzzle: The growth in actively managed mutual funds. *Journal of Finance*, 51(3):783–810, 1996. [6](#)
- Michael Hirsch, Steven Smale, and Robert Devaney. *Differential equations, dynamical systems, and an introduction to chaos*. Elsevier/Academic Press, Amsterdam, third edition, 2013. [14](#), [16](#), [18](#), [19](#)
- Julien Hugonnier and Ron Kaniel. Mutual fund portfolio choice in the presence of dynamic flows. *Mathematical Finance*, 20(2):187–227, 2010. [5](#)
- Ajay Khorana and Henri Servaes. What drives market share in the mutual fund industry? *Review of Finance*, 16:81–113, 2012. [6](#)
- Jean-Jacques Laffont and David Martimort. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press, 2009. [2](#), [8](#)
- David Luenberger. *Optimization by vector space methods*. Wiley, New York, USA, 1969. [6](#)
- Anthony Lynch and David Musto. How investors interpret past fund returns. *Journal of Finance*, 58(5):2033–2058, 2003. [5](#)
- Harry Mamaysky and Matthew Spiegel. A theory of mutual funds: optimal fund objectives and industry organization. Working paper, 2002. [5](#)
- Massimo Massa. Why so many mutual funds? Mutual fund families, market segmentation and financial performance. INSEAD Working paper, 2000. [6](#)
- Preston McAfee, John McMillan, and Michael Whinston. Multiproduct monopoly, commodity bundling, and correlation of values. *Quarterly Journal of Economics*, 104(2):371–383, 1989. [2](#), [6](#)
- Mike Mesterton-Gibbons. *A primer on the calculus of variations and Optimal control theory*, volume 50 of *Student Mathematical Library*. American Mathematical Society, Providence, Rhode Island, USA, 2009. [14](#), [29](#)
- James Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208, 1971. [13](#)

- Morningstar. Fund family 150: Due diligence on the 150 largest fund families. Morningstar due diligence reports, 2018. [1](#)
- Max Müller. Über das fundamentaltheorem in der theorie der gewöhnlichen differentialgleichungen. *Mathematische Zeitschrift*, 26(1):619–645, 1927. [18](#)
- Michael Mussa and Sherwin Rosen. Monopoly and product quality. *Journal of Economic Theory*, 18(2):301–317, August 1978. [2](#), [6](#), [8](#)
- Hui Ou-Yang. Optimal contracts in a continuous-time delegated portfolio management problem. *Review of Financial Studies*, 16(1):173–208, 2003. [5](#)
- Lubos Pastor and Robert Stambaugh. On the size of the active management industry. *Journal of Political Economy*, 120(4):740–781, 2012. [5](#)
- Jean-Charles Rochet. The taxation principle and multi-time hamilton-jacobi equations. *Journal of Mathematical Economics*, 14(2):113 – 128, 1985. [3](#)
- Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191 – 200, 1987. [13](#)
- Jean-Charles Rochet and Philippe Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826, 1998. [5](#), [13](#)
- Casey Rothschild. Nonexclusivity, linear pricing, and annuity market screening. *Journal of Risk and Insurance*, 82(1):1–32, 2015. [6](#)
- Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, Philadelphia, USA, second edition, 2014. [6](#)
- Michael Spence. Multi-product quantity-dependent prices and profitability constraints. *Review of Economic Studies*, 47(5):821–841, 1980. [2](#), [6](#)
- Lars Stole. Lectures on the theory of contracts and organizations. Lectures notes, University of Chicago, 2001. [6](#)
- Robert Wilson. *Nonlinear Pricing*. Oxford University Press, 1993. [1](#), [2](#), [6](#)