NBER WORKING PAPER SERIES

ASSET MANAGEMENT CONTRACTS AND EQUILIBRIUM PRICES

Andrea M. Buffa
Dimitri Vayanos
Paul Woolley

Working Paper 20480
http://www.nber.org/papers/w20480

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2014, Revised October 2019

We thank Ricardo Alonso, Daniel Andrei, Oliver Bought, Jennifer Carpenter, Sergey Chernenko, Chris Darnell, Peter DeMarzo, Philip Edwards, Ken French, Willie Fuchs, Diego Garcia, Jeremy Grantham, Zhiguo He, Apoorva Javadekar, Ron Kaniel, Ralph Koijen, Dong Lou, John Moore, Dmitry Orlov, Marco Pagano, Anna Pavlova, Gabor Pinter, seminar participants at the Bank of England, Banque de France, Bocconi, Boston University, CEU, Cheung Kong, Dartmouth, Duke, EIEF, Glasgow, Indiana, Insead, Louvain, LSE, Maryland, Minnesota, MIT, Naples, NY Fed, SMU, Stanford, Toulouse, UT Austin, Yale, Vienna and conference participants at AEA, ASAP, BoE Macro-Finance, BIS, CEMFI, CRETE, ESSFM Gerzenzee, FCA, FIRS, FRIC, FTG, Inquire UK, Jackson Hole, LSE PWC, NBER Asset Pricing, SFS Cavalcade, Utah, and WFA for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Asset Management Contracts and Equilibrium Prices
Andrea M. Buffa, Dimitri Vayanos, and Paul Woolley
NBER Working Paper No. 20480
September 2014, Revised October 2019
JEL No. D86,G12,G14,G18,G23

ABSTRACT

We derive equilibrium asset prices when fund managers deviate from benchmark indices to exploit noise-trader induced distortions but fund investors constrain these deviations. Because constraints force managers to buy assets that they underweight when these assets appreciate, overvalued assets have high volatility, and the risk-return relationship becomes inverted. Noise traders bias prices upward because constraints make it harder for managers to underweight overvalued assets, which have high volatility, than to overweight undervalued ones. We endogenize the constraints based on investors' uncertainty about managers' skill, and show that asset-pricing implications can be significant even for moderate numbers of unskilled managers.

Andrea M. Buffa
Department of Finance
Boston University
595 Commonwealth Avenue
Boston, MA 02215
buffa@bu.edu

Dimitri Vayanos
Department of Finance, OLD 3.41
London School of Economics
Houghton Street
London WC2A 2AE
UNITED KINGDOM
and CEPR
and also NBER
d.vayanos@lse.ac.uk

Paul Woolley
Financial Markets Group
London School of Economics
London WC2A 2AE
UNITED KINGDOM
p.k.woolley@lse.ac.uk

1 Introduction

Financial markets have become highly institutionalized. For example, individual investors were holding directly 47.9% of U.S. stocks in 1980, but only 21.5% in 2007, with most of the remainder held by financial institutions such as mutual funds and pension funds (French (2008)). The portfolios of these institutions are chosen by professional asset managers.

The institutionalization of financial markets has stimulated research on the performance of professional managers and their effects on equilibrium asset prices and market efficiency. A vast literature examines whether actively-managed funds outperform passively-managed ones. A related literature investigates whether the growth of passive funds has made markets less efficient, and whether efficiency increases in the ratio of active to passive.[1]

Drawing a sharp distinction between passive funds constrained to hold specific portfolios, and active funds investing without constraints, is in some ways misleading. This is because much of active management is done around benchmark portfolios, with managers being constrained in how much they can deviate from them. The constraints can bound a manager's tracking error (standard deviation of the difference between the manager's portfolio's return and the return of a benchmark portfolio), or the difference between the weight that the manager allocates to each asset class, geographical area, or industry sector, and the corresponding benchmark weight.[2] Viewing asset management as a continuum between active and passive, depending on the tightness of managers' constraints, seems a better description of reality. In this paper we adopt that alternative view and show that its implications for equilibrium asset prices and market efficiency differ significantly from the conventional view.

A simple example helps motivate our analysis. Suppose that some asset managers must keep their portfolio weight in each industry sector within 5% of the sector's weight in a benchmark portfolio. Suppose also that a sector that these managers view as overvalued has 10% weight in the benchmark, while the managers give it 5% weight. If the sector doubles in value, reaching 20% weight in the benchmark, then its weight in the managers' portfolio doubles to 10%, but must rise

[1] See, for example, Elton and Gruber (2013) for a survey of the literature on mutual-fund performance, and Franzoni, Ben-David, and Moussawi (2017) for a survey of exchange-traded funds (ETFs) and their effects on market performance.

[2] For a discussion of tracking-error constraints and their implications for financial markets, see the 2003 report by the Committee on the Global Financial System (BIS (2003)). According to that report, bounds on tracking error are on average 1% for actively-managed bond portfolios and between 2-6% for actively-managed stock portfolios (p.20). The Norwegian Sovereign Wealth Fund (NBIM), one of the largest institutional investors globally, reports the following regarding its tracking-error constraint: "The Ministry of Finance has set limits for how much risk NBIM may take in its active management of the fund. The most important limit is expressed as expected relative volatility (tracking error) and puts a ceiling on how much the return on the fund may be expected to deviate from the return on the benchmark portfolio. The expected tracking error limit is 125 basis points, or 1.25%." (https://www.nbim.no/en/investments/investment-risk/)

further to 15% so that the constraint is met. Buying pressure by the managers amplifies the sector's appreciation, raising its volatility. Overvalued sectors thus have high volatility, in addition to their low expected return, causing the risk-return relationship to become weak or inverted, consistent with empirical evidence.[3] Amplification does not arise when managers are constrained to hold the benchmark portfolio, or when they are fully unconstrained. The example implies additionally that overvaluation is harder to correct than undervaluation. Indeed, managers must stick closer to the benchmark in overvalued sectors: a 5% difference in weight allows less leeway in relative terms when the sector's benchmark weight is large. The links between portfolio constraints of asset managers and amplification or overvaluation have been recognized by policy-makers.[4]

In our model's basic version, described in Section 2, investors can invest in a riskless and a risky asset over an infinite horizon and continuous time. The riskless rate is constant, and the dividend flow per share of the risky asset follows a square-root process. Investors maximize a mean-variance objective over instantaneous changes in wealth. Some investors are unconstrained, while others face a constraint limiting the volatility or the dollar value of the deviation between their risky-asset position and a benchmark position. Investors may want to deviate from the benchmark position to exploit price distortions caused by noise traders.

Section 3 derives the equilibrium price of the risky asset taking the constraint as exogenous and not distinguishing between investors and the asset managers they employ. Two polar cases are analyzed first: no constraint, in which case all investors are fully active; and an infinitely tight constraint, in which case constrained investors must hold the benchmark position and hence are fully passive. In both cases, we derive a novel closed-form solution for the price and show that it is affine in the dividend flow. An increase in noise-trader demand raises the price and lowers the asset's expected return. It does not affect, however, the asset's return volatility: the price becomes more sensitive to the dividend flow, but the effect is proportional to the increase in the price level. Moving from no constraint (all investors fully active) to an infinitely tight constraint (constrained investors fully passive) exacerbates the price distortions created by noise traders. This is because the constraint prevents constrained investors from absorbing noise-trader demand. The constraint does not affect return volatility, however, because volatility is independent of demand.

Section 3 next analyzes the general case. The equilibrium involves a region where the constraint

[3]References to the empirical literature are in Section 3.4.

[4]For example, the BIS (2003) report notes: "Overvalued assets/stocks tend to find their way into major indices, which are generally capitalization-weighted and therefore will more likely include overvalued securities than undervalued securities. Asset managers may therefore need to buy these assets even if they regard them as overvalued; otherwise they risk violating agreed tracking errors." (p.19). In a similar spirit, a 2015 IMF working paper (Jones (2015)) notes: "Another source of friction capable of amplifying bubbles stems from the captive buying of securities in momentum-biased market capitalization-weighted financial benchmarks. Underlying constituents that rise most in price will see their benchmark weights increase irrespective of fundamentals, inducing additional purchases from fund managers seeking to minimize benchmark tracking error."

does not bind, and a region where it binds. The constraint binds for high values of the asset's dividend flow because the asset's price and volatility per share are high. The price in each region is characterized by a second-order ordinary differential equation (ODE), with smooth-pasting between regions. While the solution to the ODE system is not linear and not closed-form, we show analytically that it exists under general conditions and has a number of key properties.

One property of the price function is that it is convex in the dividend flow when noise-trader demand is high, and concave when demand is low. The convexity reflects the amplification effect. The concavity reflects the opposite dampening effect: since constrained investors give higher weight to an undervalued sector relative to the sector's benchmark weight, they need to sell the sector when it appreciates, dampening the appreciation. Amplification and dampening generate an inverted risk-return relationship, which we interpret as a cross-sectional one by extending our model to multiple risky assets: assets with high noise-trader demand have high volatility and low expected return, while assets with low demand have low volatility and high expected return. Moving from no constraint (all investors fully active) to intermediate levels of the constraint raises the volatility of high-demand assets and lowers that of low-demand assets. The same occurs when moving from an infinitely tight constraint (constrained investors fully passive) to an intermediate one.

A second property of the price function is that it is convex in noise-trader demand: high demand raises the price more than low demand lowers it. Intuitively, since high-demand assets have higher volatility per share than low-demand assets, investors are less willing to trade against the former assets' overvaluation than against the latter assets' undervaluation. The asymmetry arises even without the constraint, but becomes significantly more pronounced with the constraint. Because of the asymmetry, portfolios with more heterogeneous noise-trader demand across their component assets earn lower expected returns than portfolios with less heterogeneity and same average demand.

Section 4 endogenizes the constraint in a contracting model and revisits the asset-pricing analysis. We interpret the unconstrained investors as observing noise-trader demand and the dividend flow, and the constrained investors as being uninformed. Each uninformed investor can employ an asset manager, who may be skilled and observe these variables, or unskilled and observe an uninformative signal that she wrongly treats as informative. A contract consists of a fee that can depend on the investor's wealth, and a set in which the manager's deviation from a benchmark position must lie. We consider contracting over two periods, interpreting investors and managers as overlapping generations, and take the limit when the time between periods becomes small.

The optimal fee aligns the manager's risk preferences with those of the investor. The investor must guard, however, against the possibility that the manager is unskilled, and does so by restricting the manager's deviation from the benchmark position. The optimal set in which the deviation must lie is an interval ranging from zero to a positive bound. This yields the constraint assumed in Section

3. The optimal bound is infinite (no constraint) when the probability that the manager is skilled is one, and converges to zero (infinitely tight constraint) when that probability goes to zero.

The asset-pricing analysis of Section 3 carries through to an endogenous constraint. Endogenizing the constraint's parameters imposes tighter structure and allows us to express the asset-pricing effects in terms of more primitive quantities, such as the fraction of unskilled managers. In a numerical example, we find that the spread in expected return and volatility across assets with high and low noise-trader demand can be significant even for a moderate fraction of unskilled managers.

An additional advantage of endogenizing the constraint is that we can determine *effective* capital. Suppose that ten trillion dollars are invested through asset managers, and the fraction of unskilled managers is 20%. How much of the ten trillion if invested through skilled managers without any constraints would result in the same degree of market efficiency? That amount is lower than eight trillion because the presence of unskilled managers imposes constraints on the skilled ones. We find that the amount is about four trillion. Hence, abstracting away from managers' constraints can overstate significantly the available capital to correct price distortions.

Our paper relates to several strands of work on asset management and asset pricing. One literature concerns the performance of active versus passive funds, and their impact on market efficiency. That literature builds on the seminal paper by Grossman and Stiglitz (1980), in which informed and uninformed investors trade with noise traders, there is a cost to becoming informed, and price informativeness increases in the fraction of the informed. In Subrahmanyam (1991), the introduction of index futures induces noise traders to trade the index rather than the component assets. This lowers liquidity for the component assets, and has ambiguous effects on market efficiency. Related mechanisms are at play in Cong and Xu (2016) and Bhattacharya and O'Hara (2018), who study how ETFs affect market efficiency and liquidity, and Bond and Garcia (2019) who study the effects of lowering the costs of passive investing. Pastor and Stambaugh (2012) and Stambaugh (2014) explain the decline in active funds' expected returns based on the increase in the assets they manage and the decline in noise trading, respectively.[5] In Garleanu and Pedersen (2018), active funds' expected returns decline when investors are better able to locate skilled managers. In these papers, active funds invest without investor-imposed constraints, while constraints are central to our analysis.

In emphasizing constraints, our paper is related to the literature on the limits of arbitrage (see Gromb and Vayanos (2010) for a survey). In that literature, distortions are more pronounced when arbitrageurs perform poorly and become more constrained. Moreover, poor performance is generally associated with down markets (e.g., He and Krishnamurthy (2012, 2013), Brunnermeier

[5] Berk and Green (2004) show that decreasing returns to scale at the level of individual funds can explain why investors flow into funds with good past performance even though performance does not persist.

and Sannikov (2014)). By contrast, distortions in our model are more pronounced for overvalued assets and during up markets.

Another related literature studies asset management contracts. Within its strand that takes asset prices as given, our paper relates most closely to He and Xiong (2013), in which investors constrain managers' choice of assets to better incentivize them to acquire information.[6] Investors in our model constrain managers to guard against the possibility that they are unskilled. Our contracting analysis is in the spirit of the literature on optimal delegation (Alonso and Matouschek (2008), Amador and Bagwell (2013)). We draw the connections to that literature in Section 4.

Within the strand of the asset-management-contracts literature that endogenizes prices, our paper relates most closely to papers that examine the effects of compensating managers based on their performance relative to a benchmark portfolio. A common theme in several papers is that such compensation raises the price of the benchmark portfolio and of assets covarying highly with it. Brennan (1993), Basak and Pavlova (2013) and Buffa and Hodor (2018) show this result in settings where managers derive direct utility from relative performance. Kapur and Timmermann (2005) and Cuoco and Kaniel (2011) show a similar result in settings where managers receive a linear fee. The latter paper also finds that the result can reverse when the fee has option-like components. Kashyap, Kovrijnykh, Li, and Pavlova (2018) explore the result's implications for real investment.[7] Tighter constraints in our model can instead lower the price of the benchmark portfolio.

An alternative explanation for risk-return inversion is based on leverage constraints (Black (1972), Frazzini and Pedersen (2014)): investors prefer assets with high CAPM beta because they provide leverage, which investors cannot replicate by investing in low-beta assets and borrowing. Leverage constraints generate a negative relationship between CAPM beta and alpha, but a positive one between beta and expected return. In our model both relationships can be negative.

An alternative explanation for why portfolios with more heterogeneous demand across their component assets earn lower expected returns is based on short-sale constraints and disagreement between agents (Harrison and Kreps (1978), Scheinkman and Xiong (2003), Hong and Stein (2007)): short-sale constraints prevent the pessimists' views from being incorporated into the price. Our model generates a similar relationship without short-sale constraints. It also predicts that the relationship is stronger when managers' constraints are tighter.

[6] Other papers on managerial moral hazard in aquiring information include Stoughton (1993), Admati and Pfleiderer (1997), Li and Tiwari (2009), and Dybvig, Farnsworth, and Carpenter (2010). See also Bhattacharya and Pfleiderer (1985), Starks (1987), Das and Sundaram (2002), Palomino and Prat (2003), Ou-Yang (2003) and Cadenillas, Cvitanic, and Zapatero (2007) for other contracting settings.

[7] Other papers on the equilibrium effects of benchmarking include Qiu (2017) and Cvitanic and Xing (2018). See also Garcia and Vanden (2009), Gorton, He, and Huang (2010), Kyle, Ou-Yang, and Wei (2011), Malamud and Petrov (2014), Sato (2016), Huang (2018) and Sockin and Xiaolan (2018) for other models that determine jointly asset management contracts and equilibrium prices.

2 Model

Time t is continuous and goes from zero to infinity. The riskless rate is exogenous and equal to $r > 0$. A risky asset pays a dividend flow D_t per share and is in supply of θ shares. The price S_t per share of the risky asset is determined endogenously in equilibrium.

The risky asset's return per share in excess of the riskless rate is

$$dR_t^{sh} \equiv D_t dt + dS_t - rS_t dt, \tag{2.1}$$

and its return per dollar in excess of the riskless rate is

$$dR_t \equiv \frac{dR_t^{sh}}{S_t} = \frac{D_t dt + dS_t}{S_t} - rdt. \tag{2.2}$$

We refer to dR_t^{sh} as share return, omitting that it is in excess of the riskless rate. We refer to dR_t as return, omitting that it is per dollar and in excess of the riskless rate.

The dividend flow D_t follows the square-root process

$$dD_t = \kappa \left(\bar{D} - D_t\right) dt + \sigma \sqrt{D_t} dB_t, \tag{2.3}$$

where $(\kappa, \bar{D}, \sigma)$ are positive constants and B_t is a Brownian motion. The square-root specification (2.3) allows for closed-form solutions, while also ensuring that dividends remain positive. A property of the square-root specification that is key for our analysis is that the volatility (standard deviation) of dividends per share D_t increases with the level of dividends. This property is realistic: if a firm becomes larger and keeps the number of its shares constant, then its dividends per share become more uncertain in absolute terms (but not necessarily as fraction of the firm's size).[8]

Investors form a continuum with measure one. They are of two types: *unconstrained investors* who can invest in the riskless and the risky asset without any limitations, and *constrained investors* who are limited in the risk they can take. Unconstrained investors are in measure $1 - x \in (0, 1)$, and constrained investors are in the complementary measure x. We denote by W_{1t} and W_{2t} the wealth of an unconstrained and a constrained investor, respectively, and by z_{1t} and z_{2t} the number of shares of the risky asset that they hold. In Section 3 we derive equilibrium asset prices taking the constraint as exogenous. In Section 4 we endogenize the constraint in a contracting model,

[8]Dividends are often assumed to follow a geometric Brownian motion (GBM). Under the GBM specification, the volatility of dividends per share is proportional to the dividend level. Hence, the volatility of dividends per share increases with the dividend level, exactly as under the square-root specification. The two specifications have different implications for the volatility of dividends per share as *fraction* of the dividend level. Under the GBM specification that quantity is independent of the dividend level, while under the square-root specification it decreases with the dividend level. We adopt the square-root over the GBM specification because of tractability.

and revisit the equilibrium price properties. In the contracting model, constrained investors do not observe the supply θ of the risky asset, which determines the asset's expected return. They can hire an asset manager, but are uncertain about the manager's skill. Their optimal response to that uncertainty is to limit the manager's actions.

One interpretation of the assumption that constrained investors do not observe θ is that θ includes demand by noise traders, which is unobservable. That is, θ is an observable number of shares sold by the asset issuer minus an unobservable number of shares bought by noise traders. We adopt this interpretation from now on. Under this interpretation, θ can take both positive and negative values. Negative values arise when the demand by noise traders exceeds the supply by the asset issuer.

At time t, investors choose their position in the risky asset to maximize the mean-variance objective

$$\mathbb{E}_t(dW_{it}) - \frac{\rho}{2}\mathbb{V}\mathrm{ar}_t(dW_{it}), \tag{2.4}$$

subject to the budget constraint

$$dW_{it} = (W_{it} - z_{it}S_t)\, rdt + z_{it}(D_t dt + dS_t) = W_{it} rdt + z_{it} dR_t^{sh}, \tag{2.5}$$

where ρ is a risk-aversion coefficient common to all investors, $i = 1$ for unconstrained investors, and $i = 2$ for constrained investors. The mean and variance in the objective (2.4) are computed over the infinitesimal change in the investors' wealth. That change is equal to the riskless rate paid on wealth between t and $t + dt$, plus the capital gains from the risky asset in excess of the riskless rate. The capital gains are equal to the number of shares z_{it} times the share return dR_t^{sh}.

The constraint restricts the volatility of the constrained investors' position z_{2t} not to exceed a bound $L \geq 0$. We consider both the case where volatility is measured in absolute terms and the case where volatility is measured relative to a benchmark position of $\eta > 0$ shares. We nest the two cases in the constraint

$$\frac{1}{\sqrt{dt}}\sqrt{\mathbb{V}\mathrm{ar}_t\left[(z_{2t} - \eta) dR_t^{sh}\right]} \leq L, \tag{2.6}$$

where η is non-negative and becomes zero when volatility is measured in absolute terms. In Appendix B we show that our main results remain the same under the alternative constraint

$$|z_{2t} - \eta| S_t \leq L, \tag{2.7}$$

which restricts the dollar value of the constrained investors' position z_{2t} not to deviate from the dollar value of the benchmark position $\eta \geq 0$ by more than L. Equation (2.6) can be interpreted as a tracking-error constraint that restricts the volatility of portfolio return relative to a benchmark. Equation (2.7) can be interpreted as a constraint that restricts portfolio weights relative to a benchmark. We develop these interpretations in Appendix B.

Constrained investors combine elements from active and passive investing. They are active in the sense that they have some leeway when choosing their position in the risky asset. They are passive in the sense that they cannot deviate much from their benchmark. In Section 3, where we derive equilibrium asset prices with the constraint (2.6), we begin with the two polar cases where constrained investors are either fully active or fully passive. In Section 3.1 we assume $L = \infty$, which implies that there is no constraint and that constrained investors are fully active. In Section 3.2 we instead assume $L = 0$, which implies that the constraint forces constrained investors to hold η shares of the risky asset and to be fully passive. In Section 3.3 we turn to the general case where $L \in (0, \infty)$, and show that there are important qualitative differences with the two polar cases.

Investors with the objective (2.4) can be interpreted as overlapping generations living over infinitesimal periods. The generation born at time t is endowed with wealth W, invests in the riskless and the risky asset from t to $t + dt$, consumes at $t + dt$ and then dies. If preferences over consumption are described by the Von Neumann-Morgenstern (VNM) utility function U, and if all uncertainty is Brownian as is the case in equilibrium, utility maximization yields the objective (2.4) with $W_{it} = W$ and $\rho = -\frac{U''(W)}{U'(W)}$.

We endow investors with the mean-variance objective (2.4) rather than with expected utility over an infinite stream of consumption because this simplifies the equilibrium analysis of the constraint in the general case where $L \in (0, \infty)$. An additional advantage of the objective (2.4) is that it is compatible with the contracting model of Section 4, which assumes two-period contracts between overlapping generations of investors and managers, and takes the limit when the time between periods becomes small. An earlier version of this paper (Buffa, Vayanos, and Woolley (2014)) derives the equilibrium when investors have negative exponential utility over an infinite stream of consumption and are not subject to a constraint such as (2.6) or (2.7). In the polar cases studied in Sections 3.1 and 3.2, the infinite-horizon objective yields near-identical closed-form solutions and comparative statics as the mean-variance objective.

3 Equilibrium with Exogenous Constraint

3.1 No Constraint

We first derive the equilibrium when $L = \infty$. Constrained investors face no constraint and are identical to unconstrained investors. In the contracting model of Section 4, no constraint corresponds to constrained investors having access only to skilled managers.

The equilibrium price S_t is a function of the dividend flow D_t, which is the only state variable in the model. Denoting this function by $S(D_t)$ and assuming that it is twice continuously differentiable, we can write the share return dR_t^{sh} as

$$\begin{aligned} dR_t^{sh} &= D_t dt + dS(D_t) - rS(D_t)dt \\ &= \left[D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t)\right] dt + \sigma\sqrt{D_t}S'(D_t)dB_t, \end{aligned} \tag{3.1}$$

where the second step follows from (2.3) and Ito's lemma.

Using the budget constraint (2.5), we can write the objective (2.4) as

$$z_{it}\mathbb{E}_t(dR_t^{sh}) - \frac{\rho}{2}z_{it}^2\mathbb{V}\mathrm{ar}_t(dR_t^{sh}).$$

The first-order condition with respect to z_{it} is

$$\mathbb{E}_t(dR_t^{sh}) = \rho z_{it}\mathbb{V}\mathrm{ar}_t(dR_t^{sh}). \tag{3.2}$$

The expected share return $\mathbb{E}_t(dR_t^{sh})$ is the drift term in (3.1), and the share return variance $\mathbb{V}\mathrm{ar}_t(dR_t^{sh})$ is the square of the diffusion term.

Since unconstrained and constrained investors are identical, the market-clearing condition

$$(1-x)z_{1t} + xz_{2t} = \theta \tag{3.3}$$

implies $z_{1t} = z_{2t} = \theta$. Each investor's position is thus equal to the asset supply θ, which coincides with the supply per investor since investors form a continuum with mass one. Setting $z_{it} = \theta$ in (3.2), we find the following ordinary differential equation (ODE) for the function $S(D_t)$:

$$D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t) = \rho\theta\sigma^2 D_t S'(D_t)^2. \tag{3.4}$$

The ODE (3.4) is second-order and non-linear, and must be solved over $(0, \infty)$. We require that its solution $S(D_t)$ has a derivative that converges to finite limits at zero and infinity. This yields one boundary condition at zero and one at infinity.

We look for an affine solution to the ODE (3.4):

$$S(D_t) = a_0 + a_1 D_t, \tag{3.5}$$

where (a_0, a_1) are constant coefficients. This function satisfies the boundary conditions since its derivative is constant. Substituting this function into (3.4) and identifying terms, we can compute (a_0, a_1).

Proposition 3.1. *Suppose $L = \infty$ and $\theta > -\frac{(r+\kappa)^2}{4\rho\sigma^2}$. An affine solution $S(D_t) = a_0 + a_1 D_t$ to (3.4) exists, with*

$$a_0 = \frac{\kappa}{r} a_1 \bar{D}, \tag{3.6}$$

$$a_1 = \frac{2}{r + \kappa + \sqrt{(r+\kappa)^2 + 4\rho\theta\sigma^2}}. \tag{3.7}$$

Both $S(D_t)$ and $S'(D_t)$ are decreasing and convex functions of the supply θ of the risky asset.

The intuition for (3.6) and (3.7) is as follows. The coefficient a_1 is the sensitivity $S'(D_t)$ of the price to changes in the dividend flow D_t. Consider a unit increase in D_t. When the supply θ of the risky asset is equal to zero, (3.7) implies that the price S_t increases by $a_1 = \frac{1}{r+\kappa}$. This is the present value of the increase in future expected dividends discounted at the riskless rate r. Indeed, a unit increase in D_t raises the expected dividend flow $E_t(D_{t'})$ at time $t' > t$ by $e^{-\kappa(t'-t)}$. Hence, the present value of future expected dividends increases by

$$\int_t^\infty e^{-\kappa(t'-t)} e^{-r(t'-t)} dt' = \frac{1}{r+\kappa}.$$

When the supply θ of the risky asset is positive, the price S_t increases by $a_1 < \frac{1}{r+\kappa}$ in response to a unit increase in D_t. This is because the increase in D_t not only raises expected dividends, but also makes them riskier due to the square-root specification of D_t. Moreover, since investors hold a long position, the increase in risk makes them more willing to unwind their position and sell the asset. This results in a smaller price increase than when $\theta = 0$. When instead $\theta < 0$, the investors hold a short position, and the increase in risk makes them more willing to buy the asset. This results in a larger price increase than when $\theta = 0$, i.e., $a_1 > \frac{1}{r+\kappa}$. Equation (3.7) confirms that a_1 decreases in θ.

The effect of θ on a_1 is stronger when θ is small, implying that the price sensitivity $S'(D_t)$ is convex in θ. Convexity is related to a_1 being decreasing in θ and bounded below by zero. Indeed, these properties imply that the derivative of a_1 with respect to θ converges to zero when θ becomes large (while it is negative for smaller values of θ).

The coefficient a_0 is equal to the price level when the dividend flow D_t is zero. If the mean-reversion parameter κ were equal to zero, and hence the dividend flow were to stay at zero forever, then a_0 would be equal to zero. Because, however, κ is positive, and hence the dividend flow returns with certainty to positive values, a_0 is positive. Moreover, a_0 inherits properties of a_1 since the larger a_1 is, the more the price increases when the dividend flow becomes positive. In particular, a_0 is decreasing and convex in the supply θ of the risky asset, and so is the price $S_t = a_0 + a_1 D_t$. Corollary 3.1 examines how θ affects the asset's expected return and the return volatility.

Corollary 3.1. *Suppose $L = \infty$ and $\theta > -\frac{(r+\kappa)^2}{4\rho\sigma^2}$. An increase in the risky asset's supply θ raises the asset's conditional expected return $\mathbb{E}_t(dR_t)$ and leaves the return's conditional volatility $\sqrt{\mathbb{V}\mathrm{ar}_t(dR_t)}$ unaffected. The effects on the unconditional values $\mathbb{E}(dR_t)$ of expected return and $\sqrt{\mathbb{V}\mathrm{ar}(dR_t)}$ of volatility are the same as on the conditional values.*

Recall from (2.2) that the return of the risky asset is

$$dR_t = \frac{D_t}{S_t}dt + \frac{dS_t}{S_t} - rdt.$$

Return volatility is caused by the term $\frac{dS_t}{S_t}$, i.e., the capital gains per dollar invested. Since an increase in θ lowers the sensitivity a_1 of the price S_t to changes in the dividend flow D_t, it makes the capital gains $dS_t = a_1 dD_t$ per share less volatile. At the same time, the share price $S_t = a_0 + a_1 D_t$ also decreases. Because θ has the same percentage effect on a_0 and a_1, the capital gains $\frac{dS_t}{S_t}$ per dollar invested do not change, and neither does return volatility $\sqrt{\mathbb{V}\mathrm{ar}_t(dR_t)}$. On the other hand, expected return $\mathbb{E}(dR_t)$ increases because of the term $\frac{D_t}{S_t}dt$, i.e., the dividends per dollar invested. An increase in θ does not affect the dividend flow D_t per share but lowers the share price S_t.

3.2 Infinitely Tight Constraint

We next derive the equilibrium when $L = 0$. The constraint is infinitely tight and forces constrained investors to hold the benchmark position of η shares of the risky asset. In the contracting model of Section 4, an infinitely tight constraint corresponds to constrained investors having access only to unskilled managers. The benchmark position η corresponds to the position that constrained investors require their manager to hold, and is related to their expectation of θ. Hence, comparative statics with respect to θ holding η constant, such as those performed in this section, should be interpreted as changes in the unobservable component of θ, which reflects noise-trader demand.

Since the constrained investors' position z_{2t} is equal to the benchmark position η, the market-clearing condition (3.3) implies $z_{1t} = \frac{\theta - x\eta}{1-x}$. Substituting z_{1t} into (3.2) for $i = 1$, we find the ODE

$$D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t) = \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 D_t S'(D_t)^2. \tag{3.8}$$

The ODE (3.8) is identical to (3.4) except that supply θ is replaced by $\frac{\theta - x\eta}{1-x}$. The solution $S(D_t)$ of the ODE can be derived from Proposition 3.1 with the same substitution.

Proposition 3.2. *Suppose $L = 0$ and $\theta > x\eta - \frac{(1-x)(r+\kappa)^2}{4\rho\sigma^2}$. An affine solution $S(D_t) = a_0 + a_1 D_t$ to (3.8) exists, with a_0 given by (3.6) and*

$$a_1 = \frac{2}{r + \kappa + \sqrt{(r+\kappa)^2 + \frac{4\rho(\theta - x\eta)}{1-x}\sigma^2}}. \tag{3.9}$$

Relative to the case $L = \infty$:

- *$S(D_t)$ is lower when $\theta > \eta$, and higher when $\theta < \eta$.*
- *$S'(D_t)$ is lower when $\theta > \eta$, and higher when $\theta < \eta$.*

Under an infinitely tight constraint ($L = 0$), noise-trader demand has a larger effect on the price than under no constraint ($L = \infty$). Recall from Proposition 3.1 that when $L = \infty$, the price decreases in the supply θ of the risky asset. In particular, the price is higher when $\theta < \eta$, corresponding to high noise-trader demand, than when $\theta > \eta$, corresponding to low noise-trader demand. When $L = 0$, the difference is exacerbated: the price is even higher when $\theta < \eta$, and is even lower when $\theta > \eta$. Intuitively, the constraint exacerbates the effect that noise-trader demand has on the price because it prevents constrained investors from absorbing that demand. Indeed, if the constraint is imposed, constrained investors must change their position from θ to η. Therefore, they must buy the asset when $\theta < \eta$, which is when noise-trader demand is high, and must sell the asset when $\theta > \eta$, which is when noise-trader demand is low.

The constraint exacerbates the effects of noise-trader demand not only on the price level but also on the price sensitivity to changes in the dividend flow D_t. Recall from Proposition 3.1 that when $L = \infty$, the price is more sensitive to D_t (i.e., $S'(D_t)$ is larger) when $\theta < \eta$ than when $\theta > \eta$. When $L = 0$, the difference in sensitivities is exacerbated because θ is replaced by $\frac{\theta - x\eta}{1-x}$: the price becomes more sensitive to D_t when $\theta < \eta$ because $\frac{\theta - x\eta}{1-x} < \theta$, and it becomes less sensitive to D_t when $\theta > \eta$ because $\frac{\theta - x\eta}{1-x} > \theta$.

While an infinitely tight constraint exacerbates the mispricing, it does not affect return volatility. Indeed, since volatility is independent of θ, it does not change when θ is replaced by $\frac{\theta - x\eta}{1-x}$.

Corollary 3.2. *Suppose $L = 0$ and $\theta > x\eta - \frac{(1-x)(r+\kappa)^2}{4\rho\sigma^2}$. The conditional volatility $\sqrt{\mathbb{V}\mathrm{ar}_t(dR_t)}$ and the unconditional volatility $\sqrt{\mathbb{V}\mathrm{ar}(dR_t)}$ of the risky asset's return are independent of the supply θ and are the same as when $L = \infty$.*

3.3 General Case

We next derive the equilibrium for $L \in (0, \infty)$. The equilibrium is described by an unconstrained region, where the constraint does not bind, and a constrained region where it binds. Using (3.1) and assuming that $S(D_t)$ increases in D_t (which we confirm is the case in equilibrium), we can write the constraint (2.6) as

$$|z_{2t} - \eta|\sigma\sqrt{D_t}S'(D_t) \leq L. \tag{3.10}$$

In the unconstrained region all investors are identical. Therefore, their positions z_{1t} and z_{2t} are equal to the supply θ, and the function $S(D_t)$ solves the same ODE (3.4) as when the constraint never binds. Substituting $z_{2t} = \theta$ into (3.10), we find that the unconstrained region is defined by

$$|\theta - \eta|\sigma\sqrt{D_t}S'(D_t) \leq L. \tag{3.11}$$

In the constrained region (3.10) holds as an equality. Using the market-clearing condition to write z_{1t} as function of z_{2t}, and substituting into (3.2) for $i = 1$, we find

$$D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t) = \rho\frac{\theta - xz_{2t}}{1-x}\sigma^2 D_t S'(D_t)^2. \tag{3.12}$$

A binding constraint forces the position of constrained investors closer to η while keeping it on the same side of η as for unconstrained investors. When, for example, $\theta < \eta$, unconstrained investors hold a position $z_{1t} < \eta$, and constrained investors hold a position $z_{2t} \in (z_{1t}, \eta)$. Substituting z_{2t} from (3.10), which holds as an equality in the constrained region, into (3.12), and noting that $z_{2t} - \eta$ has the same sign as $\theta - \eta$, we find the ODE

$$D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t) = \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 D_t S'(D_t)^2 - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{D_t}S'(D_t), \tag{3.13}$$

where $\mathrm{sgn}(\theta - \eta)$ is the sign function, equal to one if $\theta > \eta$ and to minus one if $\theta < \eta$. The constrained region is defined by the opposite inequality to (3.11), i.e.,

$$|\theta - \eta|\sigma\sqrt{D_t}S'(D_t) > L. \tag{3.14}$$

The price function $S(D_t)$ solves the ODE (3.4) in the unconstrained region (3.11), and (3.13) in the constrained region (3.14). The two ODEs are second-order and non-linear, and must be solved as a system over $(0, \infty)$. As in Sections 3.1 and 3.2, we require that $S'(D_t)$ converges to finite limits at zero and infinity.

Since $S'(D_t)$ converges to a finite limit at zero, values of D_t close to zero belong to the unconstrained region (3.11). Conversely, since $S'(D_t)$ converges to a finite limit at infinity, values of D_t close to infinity belong to the constrained region (3.14). Hence, the unconstrained and constrained regions are separated by at least one boundary point and more generally by an odd number of such points. At a boundary point D^*, the values of $S(D^*)$ implied by the two ODEs must be equal, and the same is true for the values of $S'(D^*)$. These are the smooth-pasting conditions, and they follow from $S(D_t)$ being twice continuously differentiable. The boundary points must be solved together with the ODEs. This makes the problem a free-boundary one.

The ODE (3.4) has the affine solution derived in Proposition 3.1. That solution, however, does not satisfy the ODE (3.13), and is not the equilibrium price. While a closed-form solution to the ODEs (3.4) and (3.13) for $L \in (0, \infty)$ is not available, we can prove existence of a solution and a number of key properties.[9]

Theorem 3.1. *Suppose $L \in (0, \infty)$, $\theta > x\eta - \frac{(1-x)(r+\kappa)^2}{4\rho\sigma^2}$ and $\kappa\bar{D} > \frac{\sigma^2}{4}$. A solution $S(D_t)$ to the system of ODEs (3.4) in the unconstrained region (3.11), and (3.13) in the constrained region (3.14), with a derivative that converges to finite limits at zero and infinity, exists and has the following properties:*

- *It is increasing in D_t.*
- *It lies between the affine solution derived for $L = \infty$ and that derived for $L = 0$.*

[9] A key difficulty in proving existence is that a solution must be found over the open interval $(0, \infty)$, with a boundary condition at each end. To address this difficulty, we start with a compact interval $[\epsilon, M] \subset (0, \infty)$ and show that there exists a unique solution to the ODEs with one boundary condition at ϵ and one at M. The boundary conditions are derived from the limits of $S'(D_t)$ at zero and infinity. In the case of M, for example, the requirement that $S'(D_t)$ has a finite limit at infinity determines that limit uniquely, and we set $S'(M)$ equal to that value. To construct the solution over $[\epsilon, M]$, we use $S'(M)$ and an arbitrary value for $S''(M)$ as initial conditions for the ODEs at M, and show that there exists a unique $S''(M)$ so that the boundary condition at ϵ is satisfied. Showing uniqueness uses continuity of solutions with respect to the initial conditions, as well as a monotonicity property with respect to the initial conditions that follows from the structure of our ODEs. We next show that when ϵ converges to zero and M to infinity, the solution over $[\epsilon, M]$ converges to a solution over $(0, \infty)$. The monotonicity property of solutions with respect to the initial conditions is key to the convergence proof because it yields monotonicity of the solution with respect to ϵ and M.

- *Its derivative $S'(D_t)$ lies between the derivative of the affine solution derived for $L = \infty$ and that derived for $L = 0$.*

- *It is concave when $\theta > \eta$, and convex when $\theta < \eta$.*

- *The unconstrained and constrained regions are separated by only one boundary point D^*.*

Theorem 3.1 confirms that an increase in the dividend flow D_t raises the price S_t. It also shows that S_t lies between the values that it takes in the polar cases $L = \infty$ and $L = 0$. For given D_t, the difference in price between $\theta < \eta$ and $\theta > \eta$ is positive when there is no constraint ($L = \infty$), higher when there is a constraint ($L \in (0, \infty)$), and even higher when the constraint is infinitely tight ($L = 0$). The same comparisons hold for the difference in price sensitivity $S'(D_t)$ between $\theta < \eta$ and $\theta > \eta$.

A key difference with the polar cases $L = \infty$ and $L = 0$ is that the price is *non-linear* in D_t: it is concave for $\theta > \eta$ and convex for $\theta < \eta$, while it is affine in the polar cases. The non-linearities are driven by the trading that the constraint induces, and in turn drive the risk-return inversion. In the polar cases, there is no constraint-induced trading, either because the constraint never binds ($L = \infty$) or because constrained investors hold the benchmark position ($L = 0$).

The intuition for the non-linearities is as follows. Suppose that $\theta > \eta$ and D_t is in the constrained region. Following an increase in D_t, investors' positions go up in value and their volatility rises. To continue meeting the constraint, the constrained investors must bring their positions closer to η. Since $\theta > \eta$, they must sell some shares of the risky asset to unconstrained investors. This *dampens* the price rise. The dampening effect is weaker when D_t is smaller and in the unconstrained region because it concerns not actual sales but an expectation that sales might occur in the future. The price increase is thus larger for smaller D_t, resulting in concavity. Conversely, suppose that $\theta < \eta$ and D_t is in the constrained region. Following an increase in D_t, the constrained investors must bring their positions closer to η. Since $\theta < \eta$, they must buy some shares of the asset from unconstrained investors. This *amplifies* the price rise. The amplification effect is weaker when D_t is smaller and in the unconstrained region, resulting in convexity.

To illustrate the results shown in this and subsequent sections, we use a numerical example. We must choose values for eight parameters: the number of shares η corresponding to the benchmark position; the supply θ of the risky asset, for which we choose two values that are symmetric around η; the risk-aversion coefficient ρ of investors; the parameters $(\kappa, \bar{D}, \sigma)$ of the dividend process; the fraction x of constrained investors; and the bound L in the constraint (3.10).

We set $\rho = \eta = 1$. The choice of ρ is a normalization since we can redefine the numeraire in the units of which wealth is expressed. The choice of η is also a normalization provided that $\eta > 0$

since we can redefine one share of the risky asset by changing the dividend flow.

We set $r = 0.03$. This parameter has a small effect on our main numerical results. We set $(\kappa, \bar{D}, \sigma) = (0.05, 0.15, 0.4)$. We choose these values to generate plausible values for the expected return and return volatility in the case $\theta = \eta$. That case can be viewed as an average of the two cases $\theta > \eta$ and $\theta < \eta$ that we consider. The expected return (in excess of the riskless rate) is 5.5%, and the return volatility is 20%. These are plausible values if we interpret the risky asset as a segment of the stock market, such as mid-cap value stocks. While we use three parameters $(\kappa, \bar{D}, \sigma)$ to produce two moments, the third degree of freedom has a small effect on our numerical results.

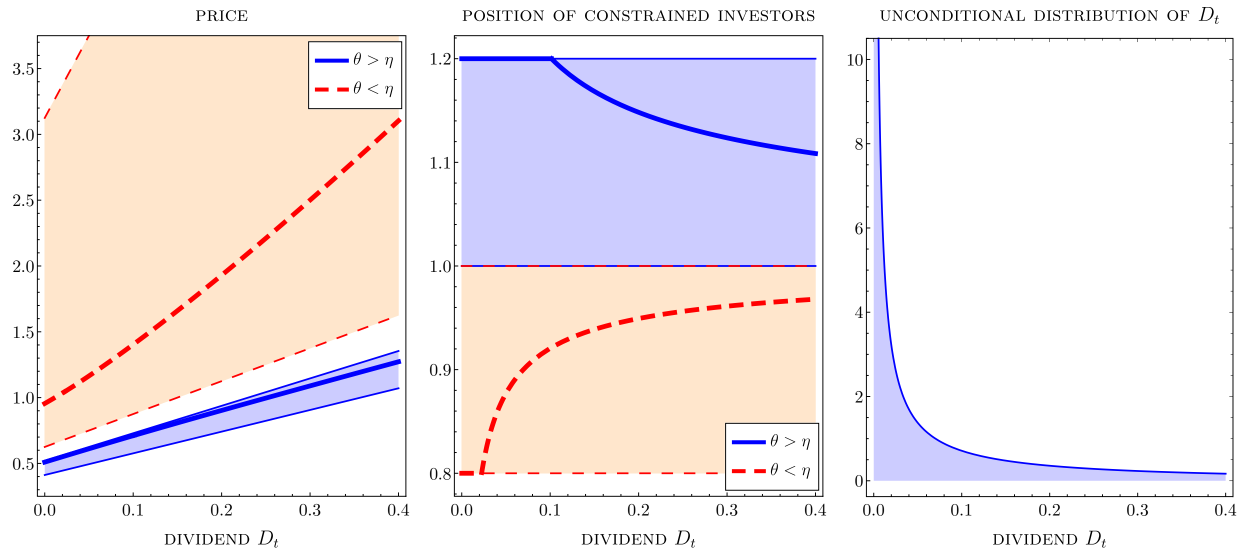
We set $x = 0.8$, i.e., 80% of investors are constrained and 20% are unconstrained. We choose the two values of θ to be 0.8 and 1.2. Thus, if the asset issuer sells 1.2 shares, then noise traders hold no shares when $\theta = 1.2$, and hold 0.4 shares, i.e., one-third of the issued quantity, when $\theta = 0.8$. Our numerical results become stronger when x increases or when the spread $\Delta\theta$ between the high and the low value of θ increases. We finally set $L = 0.05$. In subsequent sections we examine the effects of varying $\Delta\theta$ and L.

Figure 1 illustrates the results of Theorem 3.1 in our numerical example. The left panel plots the price as function of D_t. The thick lines represent the price when there is a constraint ($L \in (0, \infty)$). The thin lines represent the price in the two polar cases where there is no constraint ($L = \infty$) and where the constraint is infinitely tight ($L = 0$), with the price in the latter case corresponding to the more extreme values. In all three cases, the dashed red line is drawn for $\theta = 0.8$, corresponding to high noise-trader demand, and the solid blue line is drawn for $\theta = 1.2$, corresponding to low noise-trader demand. The area between the price in the two polar cases is shaded. Consistent with Theorem 3.1, the thick lines lie inside the shaded area. The middle panel of Figure 1 plots the position of constrained investors using the same conventions. The position when $L = 0$ coincides with the horizontal line $z_{2t} = 1$. The right panel of Figure 1 plots the unconditional distribution of D_t. Besides confirming the properties shown in Theorem 3.1, Figure 1 shows that the constraint has a larger effect on prices and positions when $\theta = 0.8$ than when $\theta = 1.2$. We return to this asymmetry in subsequent sections.

3.4 Risk-Return Inversion

Unlike in the polar cases $L = \infty$ and $L = 0$, where supply θ does not affect the asset's return volatility, volatility is affected by θ when $L \in (0, \infty)$. Volatility is higher when $\theta < \eta$, corresponding to high noise-trader demand, than when $\theta > \eta$, corresponding to low noise-trader demand.

Figure 1: Effect of Constraint on Prices and Positions



Asset price S_t (left panel) and position z_{2t} of constrained investors (middle panel) as functions of the dividend flow D_t, and unconditional distribution of D_t (right panel). The thick lines in the left and middle panels correspond to the case where there is a constraint ($L \in (0, \infty)$). The thin lines correspond to the polar cases where there is no constraint ($L = \infty$) and where the constraint is infinitely tight ($L = 0$). The latter case corresponds to the more extreme values in the left panel and to the horizontal line $z_{2t} = 1$ in the middle panel. In all three cases, the dashed red line is drawn for $\theta = 0.8$ and the solid blue line is drawn for $\theta = 1.2$. Parameter values are: $\eta = \rho = 1$, $r = 0.03$, $\kappa = 0.05$, $\bar{D} = 0.15$, $\sigma = 0.4$, $x = 0.8$, $L = 0.05$.

Proposition 3.3. *Under the assumptions in Theorem 3.1, both the conditional volatility* $\sqrt{\mathbb{V}\mathrm{ar}_t(dR_t)}$ *and the unconditional volatility* $\sqrt{\mathbb{V}\mathrm{ar}(dR_t)}$ *of the asset's return are:*

- *Higher when* $\theta < \eta$ *than when* $\theta > \eta$*.*
- *Higher than under the affine solutions derived for* $L = \infty$ *and for* $L = 0$ *when* $\theta < \eta$*, and lower when* $\theta > \eta$*.*

Proposition 3.3 is related to the convexity and concavity results of Theorem 3.1. The amplification effect that generates the convexity for $\theta < \eta$ also generates the high volatility. The dampening effect that generates the concavity for $\theta > \eta$ also generates the low volatility.

Proposition 3.3 implies a negative relationship between volatility and expected return. When $\theta < \eta$, expected return is low, so that investors are induced to hold small positions, and volatility is high. When instead $\theta > \eta$, expected return is high, so that investors are induced to hold large positions, and volatility is low. High volatility goes together with overvaluation (low expected return) because they are both driven by high noise-trader demand. Indeed, to accommodate the high demand, investors underweight the asset relative to the benchmark position η. When the market goes up, the constraint forces them to underweight less, and hence to buy the asset. This yields amplification and high volatility.

A negative relationship between volatility and expected return has been documented empirically, and is known as the *volatility anomaly* because it is at odds with standard theories. Haugen and Baker (1996) and Ang, Hodrick, Xing, and Zhang (2006) document the volatility anomaly in the cross-section of U.S. stocks.

The volatility anomaly in our model holds as a comparative-statics rather than as a cross-sectional result because there is only one risky asset. We can, however, extend our model to multiple risky assets and derive a cross-sectional result. The simplest way to perform the extension is to assume that dividend flows are independent across assets, and that the constraint applies asset-by-asset rather than across an entire portfolio.

The multi-asset version of our model is as follows. There are N risky assets instead of one. Asset $n = 1, .., N$ pays a dividend flow D_{nt} per share and is in supply of θ_n shares. The dividend flow D_{nt} follows the square-root process

$$dD_{nt} = \kappa_n \left(\bar{D}_n - D_{nt}\right) dt + \sigma_n \sqrt{D_{nt}} dB_{nt}, \tag{3.15}$$

which generalizes (2.3), and the Brownian motions $\{dB_{nt}\}_{n=1,..,N}$ are independent. Investors can invest in all N risky assets. Generalizing (3.10), we require that the volatility of the constrained investors' position z_{2nt} in risky asset n does not exceed an upper bound

$$|z_{2nt} - \eta| \sigma_n \sqrt{D_{nt}} S'(D_{nt}) \leq L. \tag{3.16}$$

The multi-asset version of our model is a replica of the one-asset version: the price of asset n is given by Proposition 3.1 when $L = \infty$, by Proposition 3.2 when $L = 0$, and by Theorem 3.1 when $L \in (0, \infty)$. Proposition 3.3 holds as a cross-sectional comparison between assets (n, n') with $\theta_n < \eta$, $\theta_{n'} > \eta$ and identical other characteristics.

The negative relationship between volatility and expected return holds for assets that differ in θ. One could alternatively compare, in the spirit of standard theories, assets that differ in the dividend volatility coefficient σ. That comparison yields a positive relationship: assets with high σ have higher volatility and expected return than assets with lower σ and identical other characteristics. The positive relationship weakens when θ varies, and becomes negative when variation in θ dominates variation in σ.

Figure 2 illustrates risk-return inversion in our numerical example. The left and middle panels plot expected return and return volatility, respectively, as function of L. In both panels, the dashed red line is drawn for an asset n with $\theta_n = 0.8$, corresponding to high noise-trader demand, and the solid blue line is drawn for an asset n' with $\theta_{n'} = 1.2$, corresponding to low noise-trader demand. Consistent with Propositions 3.1 and 3.2, the difference in expected returns between $\theta_n = 0.8$ and $\theta_{n'} = 1.2$ increases when the constraint tightens (L decreases). The increase is from approximately 2% when L exceeds 0.15 to 7% when L is close to zero, with the effect driven primarily by the asset with high noise-trader demand ($\theta_n = 0.8$). Consistent with Proposition 3.3, the difference in return volatilities between $\theta_n = 0.8$ and $\theta_{n'} = 1.2$ is largest for intermediate values of L. When L is close to zero or exceeds 0.15, volatility is approximately 20%, for both $\theta_n = 0.8$ and $\theta_{n'} = 1.2$. When instead $L = 0.05$, volatility rises to approximately 24% for $\theta_n = 0.8$, and drops to 19% for $\theta_{n'} = 1.2$. The difference in volatilities for intermediate values of L is thus driven primarily by the amplification effect for the asset with high noise-trader demand ($\theta_n = 0.8$). This asymmetry, and the corresponding one for expected return, parallel the asymmetry shown in Figure 1.

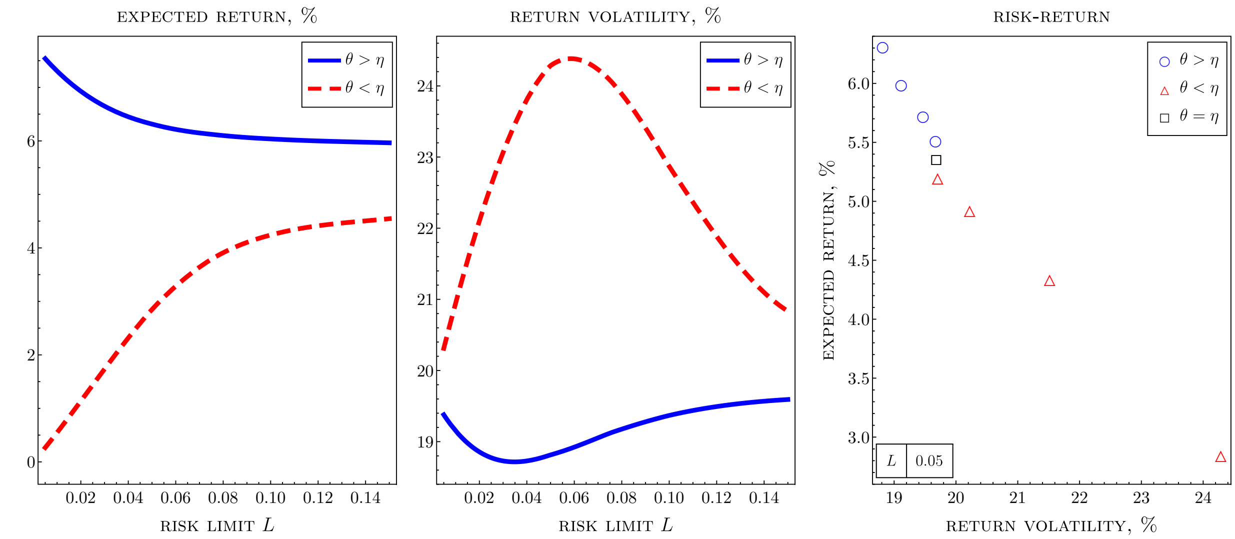
The right panel of Figure 2 plots expected return as function of return volatility for $L = 0.05$. Each risky asset corresponds to one point. We include the asset pair corresponding to $\theta_n = 0.8$ and $\theta_{n'} = 1.2$, as well as additional pairs obtained by varying the spread $\Delta\theta$ between the high and the low value of θ_n while holding the average of the two values equal to $\eta = 1$. The red triangles correspond to assets with $\theta_n < \eta = 1$, the blue circles to assets with $\theta_n > \eta = 1$, and the black square to an asset with $\theta_n = 1$. Consistent with Proposition 3.3, variation driven by θ_n generates a negative relationship between volatility and expected return.

An additional measure of risk that we can relate to expected return is CAPM beta. The CAPM predicts a positive relationship between beta and expected return. Empirically, however, a flat or negative relationship has been documented, and is known as the *beta anomaly*. Black (1972), Black, Jensen, and Scholes (1972), and Frazzini and Pedersen (2014) document a flat relationship in the cross-section of U.S. stocks. Baker, Bradley, and Wurgler (2011) find that the relationship turns negative in recent decades.

The multi-asset version of our model yields a negative relationship between beta and expected return. This is because with independent dividend flows, an asset's beta is proportional to the asset's return variance times the asset's weight in the market portfolio. Assets with $\theta_n < \eta$ have high beta because they have both high return volatility (Proposition 3.3) and high market-portfolio weight because of their high price (Theorem 3.1). Note that the negative relationship between beta and expected return arises even in the polar cases $L = \infty$ and $L = 0$. In these cases, the return volatility is independent of θ_n (Corollaries 3.1 and 3.2), but the price is higher for low-θ_n assets (Propositions 3.1 and 3.2).

Proposition 3.4. *In the multi-asset version of our model, suppose* $\theta_n > x\eta - \frac{(1-x)(r+\kappa_n)^2}{4\rho\sigma_n^2}$ *and*

Figure 2: Risk-Return Inversion



Expected return $\frac{1}{dt}\mathbb{E}(dR_t)$ and return volatility $\frac{1}{\sqrt{dt}}\sqrt{\mathbb{V}\text{ar}(dR_t)}$ as functions of L (left and middle panels), and expected return as function of return volatility for fixed L (right panel). Moments are unconditional. In the left and middle panels, the dashed red line is drawn for an asset n with $\theta_n = 0.8$ and the solid blue line is drawn for an asset n' with $\theta_{n'} = 1.2$. In the right panel, the black square corresponds to an asset with $\theta_n = 1$, the red triangles to assets with $\theta_n < \eta = 1$ and the blue circles to assets with $\theta_n > \eta = 1$, where $\theta_n = 1 \pm \frac{\Delta\theta}{2}$ and $\Delta\theta \in \{0, 0.1, 0.2, 0.3, 0.4\}$. The remaining parameter values are as in Figure 1.

$\kappa_n \bar{D}_n > \frac{\sigma_n^2}{4}$ for all $n = 1, .., N$, and consider assets (n, n') with $\theta_n < \eta < \theta_{n'}$ and identical other characteristics. Asset n has higher conditional and unconditional CAPM beta than asset n'.

3.5 Overvaluation Bias

In this section we show that the effects of noise-trader demand do not cancel out when aggregating assets into portfolios. We assume that the asset market consists of segments and that each segment consists of sub-segments. For example, one segment could be mid-cap stocks, and its sub-segments could be mid-cap value and mid-cap growth stocks. We identify the assets in our model with the sub-segments, and assume that noise-trader demand differs across them.

Using Propositions 3.1 and 3.2, we can determine the effect of aggregation in the polar cases $L = \infty$ and $L = 0$. The propositions show that the price is a convex function of θ. Hence, a segment in which θ varies across sub-segments trades at a higher price than a segment with no such variation and same average θ. Noise-trader demand thus does not cancel out by aggregation, but introduces an *overvaluation bias*. Proposition 3.5 shows overvaluation bias in the polar cases $L = \infty$ and $L = 0$, assuming for simplicity that each segment consists of two assets.

Proposition 3.5. *In the multi-asset version of our model, suppose $L = \infty$ or $L = 0$, and $\theta_n > x\eta - \frac{(1-x)(r+\kappa_n)^2}{4\rho\sigma_n^2}$ for all $n = 1,..,N$. For a segment consisting of assets (n, n') and a segment consisting of assets $(\hat{n}, \hat{n}')$ with $\theta_n < \theta_{\hat{n}} \leq \theta_{\hat{n}'} < \theta_{n'}$, $\frac{\theta_n+\theta_{n'}}{2} = \frac{\theta_{\hat{n}}+\theta_{\hat{n}'}}{2} \equiv \bar{\theta}$, and other characteristics being identical across assets,*

$$\mathcal{O}(D_t) \equiv [S_n(D_t) + S_{n'}(D_t)] - [S_{\hat{n}}(D_t) + S_{\hat{n}'}(D_t)] > 0. \tag{3.17}$$

Moreover, $\mathcal{O}(D_t)$ is larger when $L = 0$ than when $L = \infty$ under the sufficient condition $\bar{\theta} \leq \eta$.

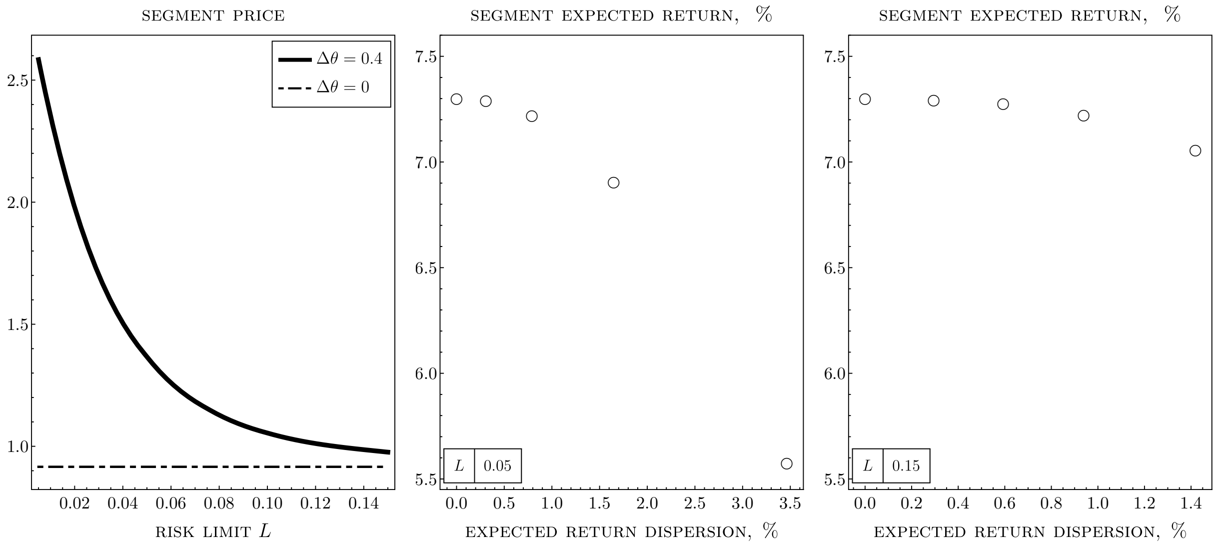
Proposition 3.5 implies a negative relationship between the variability of noise-trader demand, or equivalently of expected returns, within a segment, and the segment's own expected return. Intuitively, the negative relationship arises because the price sensitivity $S'(D_t)$ to the dividend flow D_t decreases in θ (Proposition 3.1). When θ is large, volatility per share is low because price sensitivity is low. Hence, an increase in the number of shares θ causes a small price drop. When instead θ is small, volatility per share is high, and hence an equal decrease in θ causes a large price rise. Averaging across the two cases, a segment with more extreme values of θ trades at a higher price than a segment with less extreme values.

The negative relationship between within-segment variability of noise-trader demand and segment expected return arises even when there is no constraint ($L = \infty$). Proposition 3.5 shows that the relationship becomes stronger when the constraint tightens (from $L = \infty$ to $L = 0$). This is because the constraint prevents constrained investors from absorbing noise-trader demand, increasing the demand's effective variability.

Figure 3 illustrates overvaluation bias in our numerical example. The left panel plots the unconditional averages of the prices of two segments as function of L: the segment with $(\theta_n, \theta_{n'}) = (0.8, 1.2)$, represented by the thick line, and the segment with $(\theta_{\hat{n}}, \theta_{\hat{n}'}) = (1, 1)$, represented by the thin line. Consistent with Proposition 3.5, the former segment trades at a higher price, and the price difference increases when L decreases. Since the price of the latter segment does not depend on L (the constraint does not bind for assets $(\hat{n}, \hat{n}')$ because $\theta_{\hat{n}} = \theta_{\hat{n}'} = \eta = 1$), the price of the former segment increases when L decreases. This reflects the asymmetry shown in Figure 1: the constraint raises the price of the asset with $\theta_n = 0.8$ more than it lowers the price of the asset with $\theta_{n'} = 1.2$.

The middle and right panels of Figure 3 plot expected return at the segment level as function of the dispersion in expected returns within the segment. Each segment corresponds to one point. We include the segment with $(\theta_n, \theta_{n'}) = (0.8, 1.2)$, as well as additional segments obtained by varying the spread $\Delta\theta$ between the high and the low value of θ while holding the average of the two values

Figure 3: Overvaluation Bias



Segment price $\frac{1}{2}\mathbb{E}\left(S_{nt} + S_{n't}\right)$ (left panel) as function of L, and segment expected return $\frac{1}{dt}\mathbb{E}\left(\frac{(D_{nt}+D_{n't})dt+d(S_{nt}+S_{n't})}{S_{nt}+S_{n't}}\right) - r$ as function of the dispersion in expected returns $\frac{1}{dt}\left(\mathbb{E}(dR_{n't}) - \mathbb{E}(dR_{nt})\right)$ within the segment, for fixed L (middle and right panels). Prices and expected returns are unconditional. Segments are asset pairs (n, n') with $\theta_n = 1 - \frac{\Delta\theta}{2}$ and $\theta_{n'} = 1 + \frac{\Delta\theta}{2}$, where $\Delta\theta \in \{0, 0.4\}$ in the left panel and $\Delta\theta \in \{0, 0.1, 0.2, 0.3, 0.4\}$ in the middle and right panels. The remaining parameter values are as in Figure 1.

equal to $\eta = 1$. The middle panel is drawn for $L = 0.05$, and the right panel for $L = 0.15$. Both panels show a negative relationship between within-segment dispersion in expected returns and segment expected return. The slope of the relationship is steeper (more negative) when $L = 0.05$, consistent with the comparison that Proposition 3.5 derives between $L = \infty$ and $L = 0$.

4 Equilibrium with Endogenous Constraint

4.1 Contracts

We begin with a two-period contracting model with exogenous asset returns. In Section 4.2 we embed that model into the continuous-time equilibrium model of Section 3. There are two periods, 0 and 1. The riskless rate is exogenous and equal to r. A risky asset has return R in excess of the riskless rate.

There are two agents, an investor (he) and an asset manager (she). The investor has a prior distribution Π_0 for R. The manager is either skilled or unskilled. A skilled manager forms a posterior distribution $\Pi(s)$ for R based on an informative signal s that she observes. An unskilled manager observes an uninformative signal but she wrongly treats it as informative. Her signal makes her either optimistic, with posterior distribution Π^O, or pessimistic, with posterior distribution Π^P. These two outcomes are equally likely. The probability that the manager is unskilled is $\lambda \in [0, 1)$. The investor and the manager have negative exponential utility over consumption in period one, with coefficient of absolute risk aversion equal to ρ for the investor and $\bar{\rho}$ for the manager. Without loss of generality, we set the investor's initial wealth to zero.

The investor offers the manager a contract in period 0. If the manager accepts the contract, then she observes her private signal and chooses a portfolio for the investor. The portfolio consists of z shares in the risky asset and $-zS$ dollars in the riskless asset, where S is the risky asset's price in period 0. The investor's wealth W in period 1 is $W = zSR$, equal to the dollars zS invested in the risky asset times that asset's return R in excess of the riskless asset.

The signal s observed by the skilled manager is continuous and takes values in a set Φ with probability density $h(s)$. The posterior distribution $\Pi(s)$ gives positive probability to positive and to negative values of R. As a consequence, the position z_0^* in the risky asset that maximizes the investor's unconditional expected utility, and the position $z^*(s)$ that maximizes his expected utility conditional on s, are finite. We take the range of $z^*(s)$ to be the real line. This is without loss of generality since we can introduce additional signals that have arbitrarily small probability and fill any gaps in $z^*(s)$.

In contrast to the skilled manager, the unskilled manager gives positive probability only to positive values of R under the optimistic posterior distribution Π^O, and only to negative values under the pessimistic distribution Π^P. Thus, the unskilled manager either believes that the risky asset has no downside or that it has no upside. Our analysis extends to more moderate beliefs by the unskilled manager when parametric restrictions are imposed on the asset return distribution.[10]

The contract consists of a fee, which the investor pays to the manager out of his period 1 wealth W, and of an investment restriction. The fee can be a general function $f(W)$, subject to a non-negativity and a monotonicity constraint. The non-negativity constraint is $f(W) \geq 0$, and arises because the manager has limited liability. The monotonicity constraint is that $f(W)$ is increasing, and could arise from moral hazard in period 1. Indeed, a decreasing fee could incentivize the manager to engage in wasteful activities that reduce W so to increase her fee. A non-decreasing fee could also incentivize such activities if they yield an infinitesimally small private benefit to the manager. An additional reason to assume an increasing fee is to rule out the implausible outcome that the investor can induce the manager to choose any position Z just by offering her a constant fee

[10] See, for example, Vayanos (2018), where the asset return can take two values, the skilled and the unskilled manager give positive probability to each value, and the unskilled manager has more extreme beliefs than the skilled manager. Beliefs are more extreme in the sense that the unskilled manager's probability of the high value either exceeds the skilled manager's maximum probability across all realizations of s, or is lower than the minimum probability.

and exploiting her indifference. To ensure that an optimal fee exists, we formulate the monotonicity constraint as a weak rather than a strict inequality: $f'(W) \geq \epsilon g'(W) > 0$ where ϵ is a positive constant and $g(W)$ is an increasing and bounded function defined over $(-\infty, \infty)$. We derive the optimal fee for each ϵ, and take the limit when ϵ goes to zero.

The investment restriction concerns the position z chosen by the manager. In period 0, the investor observes a statistic $F(\Delta, s)$ that depends on $\Delta \equiv |z - \eta| S$, the dollar deviation between the manager's position z and a benchmark position η, and can also depend on the signal s. For example, the investor could observe the deviation Δ, or the portfolio return volatility $\sqrt{\mathbb{V}\text{ar}_s[\Delta R]} = \Delta\sqrt{\mathbb{V}\text{ar}_s(R)}$ relative to the benchmark position. We assume that the statistic $F(\Delta, s)$ is non-negative, is continuously differentiable in Δ, and satisfies $F(0, s) = 0$ and $\frac{\partial F(\Delta, s)}{\partial \Delta} > 0$, as is the case in the above examples. The investor can choose the benchmark position η and can restrict the statistic $F(\Delta, s)$ to lie in a closed set $\mathcal{L}$.

A fee $f(W)$, benchmark position η and set $\mathcal{L}$ constitute a feasible contract. The investor chooses such a contract to maximize his expected utility. He is subject to the incentive-compatibility (IC) constraint on the manager's choice of position Z. He must also ensure that the fee satisfies non-negativity and monotonicity. Non-negativity ensures that the manager's individual rationality (IR) constraint is satisfied.

Our contracting model is in the spirit of the literature on optimal delegation (e.g., Alonso and Matouschek (2008), Amador and Bagwell (2013)). A key result in that literature is that instead of taking an action based on information sent by the agent, the principal can equivalently let the agent take the action within a restricted *delegation set*. The delegation literature generally precludes monetary transfers between the principal and the agent. We allow monetary transfers, but in the spirit of the delegation literature, restrict the fee function $f(W)$ to not depend on information sent by the agent. We also restrict the delegation set $\mathcal{L}$ to depend only on the statistic $F(\Delta, s)$ that the principal observes. The restriction on $\mathcal{L}$ could be arising from investors' limited ability to process information. For example, investors could observe that their managers deviated significantly from a benchmark, but might be unable to assess whether the benchmark's volatility was high to warrant such a deviation.

Proposition 4.1 determines the optimal fee $f(W)$, benchmark position η and delegation set $\mathcal{L}$ in the limit when ϵ goes to zero. The set $\mathcal{L}$ takes the form $[0, L]$, with $L > 0$. We denote by $\bar{z}(s)$ and $\underline{z}(s)$, respectively, the maximum and minimum value of z that meet the constraint $F(\Delta, s) \leq L$. These values are well-defined because $F(\Delta, s)$ is continuous and increasing in Δ. We denote by $\bar{\Phi}$ the set of signals s such that the investor's optimal position $z^*(s)$ conditional on s exceeds $\bar{z}(s)$, and by $\underline{\Phi}$ the set of signals s such that $z^*(s) < \underline{z}(s)$.

Proposition 4.1. *In the limit when ϵ goes to zero:*

- *The investor employs the manager.*
- *The optimal delegation set $\mathcal{L}$ has the form $[0, L]$, with $L > 0$.*
- *The position $z_S(s)$ chosen by the skilled manager is $z^*(s)$ when $s \in \Phi \backslash (\bar{\Phi} \cup \underline{\Phi})$, $\bar{z}(s)$ when $s \in \bar{\Phi}$, and $\underline{z}(s)$ when $s \in \underline{\Phi}$.*
- *The position $z_U(s)$ chosen by the unskilled manager is $\bar{z}(s)$ when her posterior is Π^O, and $\underline{z}(s)$ when her posterior is Π^P.*
- *The optimal benchmark position η and bound L solve*

$$\max_{\eta, L} \left\{ -(1-\lambda) \int_{s \in \Phi} \mathbb{E}_s \left(e^{-\rho z_S(s) SR} \right) h(s) ds - \lambda \int_{s \in \Phi} \mathbb{E}_0 \left(e^{-\rho z_U(s) SR} \right) h(s) ds \right\}. \tag{4.1}$$

- *The optimal fee $f(W)$ converges to zero for all W.*

The optimal delegation set has the same form as under the constraint in Section 2: the investor restricts the statistic $F(\Delta, s)$ not to exceed a bound L. The investor does not exclude values in the interior of $[0, L]$ because they arise only from positions chosen by the skilled manager. Indeed, since the fee is increasing, the unskilled manager always chooses extreme positions, which render $F(\Delta, s)$ equal to L.

The optimal fee $f(W)$ aligns the manager's risk preferences with those of the investor. It ensures, in particular, that the position maximizing the skilled manager's expected utility given her signal s coincides with the investor's optimal position $z^*(s)$. The position $z^*(s)$ is chosen if it gives rise to a value of $F(\Delta, s)$ that does not exceed L. If instead $F(\Delta, s)$ exceeds L, then the chosen position is the one that renders $F(\Delta, s)$ equal to L and is closest to $z^*(s)$. Proposition 4.1 shows that a fee converging to zero when ϵ goes to zero suffices to align risk preferences.

The optimal bound L trades off the cost of restricting the position chosen by the skilled manager when her signal calls for a large position (i.e., $z^*(s)$ gives rise to a value of $F(\Delta, s)$ that exceeds L), and the benefit of restricting the position chosen by the unskilled manager. These cost and benefit correspond to the first and second term, respectively, in (4.1).

The investor always employs the manager despite the risk that she may be unskilled. Indeed, when not employing the manager, the investor chooses z_0^*, his optimal position given the prior distribution Π_0. He can replicate that outcome by employing the manager and setting $(\eta, L) = (z_0^*, 0)$. He can also do strictly better by raising L slightly, giving the manager some discretion. Indeed, deviations by the unskilled manager generate a second-order loss for the investor because

z_0^* is optimal. By contrast, deviations by the skilled manager generate a first-order gain because they occur when z_0^* is not optimal and bring the investor closer to his optimal position $z^*(s)$. The manager's fee is an additional cost for the investor, but it converges to zero when ϵ goes to zero.

4.2 Equilibrium

We next embed the contracting model of Section 4.1 into the equilibrium model of Section 3. We interpret investors as overlapping generations living over infinitesimal periods. A measure $1-x$ of investors of each generation observe the asset supply θ and the dividend flow D_t. They correspond to the unconstrained investors of Section 3. The complementary measure x of investors do not observe (θ, D_t) and can employ a manager. They correspond to the constrained investors. Managers can be skilled or unskilled, as described in Section 4.1. Skilled managers observe (θ, D_t). We set $s \equiv (\theta, D_t)$ and take the statistic $F(\Delta, s)$ to be the portfolio return volatility

$$F(\Delta, s) = \frac{1}{\sqrt{dt}}\sqrt{\mathbb{V}\mathrm{ar}_s\left[(z_{2t}-\eta)S_t dR_t\right]} = \frac{1}{\sqrt{dt}}\sqrt{\mathbb{V}\mathrm{ar}_s\left[(z_{2t}-\eta)dR_t^{sh}\right]} = |z_{2t}-\eta|\sigma\sqrt{D_t}S'(D_t).$$

Solving for equilibrium with an endogenous constraint involves a fixed-point problem: asset prices must clear the market given the constraint, and the constraint (i.e., the delegation set) must be optimal given equilibrium prices. The determination of equilibrium prices given the constraint is along the same lines as in Section 3. The only change is that the uninformed investors whose manager turns out to be unskilled do not invest optimally subject to the constraint. Half of them employ an optimistic manager, who invests the maximum value of z_{2t} that meets the constraint. The remaining half employ a pessimistic manager, who invests the minimum value. Since the average of the maximum and the minimum value is η, and the measure of uninformed investors employing an unskilled manager is λx, the market-clearing condition (3.3) is replaced by

$$(1-x)z_{1t} + (1-\lambda)xz_{2t} + \lambda x\eta = \theta. \tag{4.2}$$

The definition of the unconstrained and the constrained regions is modified similarly. Since in the unconstrained region $z_{1t} = z_{2t}$, (4.2) implies $z_{2t} = \frac{\theta-\lambda x\eta}{1-\lambda x}$. Substituting into the constraint $F(\Delta, s) \leq L$, we find that the unconstrained region is defined by

$$\frac{|\theta-\eta|}{1-\lambda x}\sigma\sqrt{D_t}S'(D_t) \leq L, \tag{4.3}$$

which replaces (3.11). The ODE system is modified similarly, as shown in the proof of Proposition 4.2, and the existence and characterization results of Theorem 3.1 hold.

The determination of the constraint given equilibrium prices follows Proposition 4.1. The optimal values of (η, L) maximize the continuous-time limit of the objective (4.1). Proposition 4.2 characterizes the solution to this problem.

Proposition 4.2. *Suppose that θ takes values in a discrete set with probabilities $\pi(\theta)$. The first-order conditions with respect to η and L are*

$$\sum_{\theta>\eta} \pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\bar{z}(s)]1_{\{z^*(s)>\bar{z}(s)\}}\sigma^2 D_t S'(D_t)^2\right] + \sum_{\theta<\eta} \pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\underline{z}(s)]1_{\{z^*(s)<\underline{z}(s)\}}\sigma^2 D_t S'(D_t)^2\right]$$
$$+\frac{\lambda}{1-\lambda}\sum_{\theta}\pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\eta]\sigma^2 D_t S'(D_t)^2\right]=0, \tag{4.4}$$
$$\sum_{\theta>\eta} \pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\bar{z}(s)]1_{\{z^*(s)>\bar{z}(s)\}}\sigma\sqrt{D_t}S'(D_t)\right] - \sum_{\theta<\eta} \pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\underline{z}(s)]1_{\{z^*(s)<\underline{z}(s)\}}\sigma\sqrt{D_t}S'(D_t)\right]$$
$$-\frac{\lambda}{1-\lambda}L=0, \tag{4.5}$$

respectively, where

$$\bar{z}(s)=\eta+\frac{L}{\sigma\sqrt{D_t}S'(D_t)},$$
$$\underline{z}(s)=\eta-\frac{L}{\sigma\sqrt{D_t}S'(D_t)},$$
$$z^*(s)=\begin{cases}\frac{\theta-\lambda x\eta}{1-\lambda x} & \text{if } \frac{|\theta-\eta|}{1-\lambda x}\sigma\sqrt{D_t}S'(D_t)\le L,\\ \frac{\theta-x\eta-\frac{sgn(\theta-\eta)(1-\lambda)xL}{\sigma\sqrt{D_t}S'(D_t)}}{1-x} & \text{if } \frac{|\theta-\eta|}{1-\lambda x}\sigma\sqrt{D_t}S'(D_t)> L.\end{cases}$$

The optimal benchmark η and bound L have the following properties:

- *When θ can take only one value, $\eta=\theta$ and $L=0$.*
- *When θ can take multiple values, $\eta\in(\theta_{min},\theta_{max})$, where θ_{min} is the minimum and θ_{max} is the maximum value of θ. Moreover, $L=\infty$ when $\lambda=0$, $L\in(0,\infty)$ when $\lambda>0$, and $\lim_{\lambda\to 1}L=0$.*

When θ can take only one value, the uninformed investors set the benchmark position η equal to that value. They also set the bound L in the constraint to zero, hence requiring the manager to hold θ shares of the risky asset. This is an optimal arrangement since θ is the asset supply and hence the position that each investor in the mass-one continuum should be holding in equilibrium.

When θ can take multiple values, uninformed investors are generally unable to replicate the above arrangement because they do not observe θ. The only case where replication is possible is when all managers are skilled ($\lambda=0$). In that case, uninformed investors set L to infinity, knowing that skilled managers will choose θ. When instead some managers are unskilled, investors set L to a finite value. In the limit where all managers are unskilled, L goes to zero.

The optimal value of η is related to the uninformed investors' expectation of θ. It is not equal to that expectation, however, but to a weighted average in which smaller values of θ receive larger weight. Indeed, since volatility per share is higher for small θ, a constraint in which η is equal to the expectation of θ would restrict managers more tightly for small θ: it would bind more often for a position smaller than η, than for a position that deviates from η by an equal number of shares in the opposite direction. Moreover, the former deviation is more profitable than the latter, and hence should be less restricted. Indeed, since high noise-trader demand raises the price more than low demand lowers it (Section 3.5), price distortions are more pronounced for small θ.

Figures 1-3 illustrate the asymmetry that causes the optimal η to be smaller than the investors' expectation of θ. Suppose that each of $\theta = 0.8$ and $\theta = 1.2$ has probability 0.5, in which case the investors' expectation of θ is equal to one, the value of η assumed in the figures. The middle panel of Figure 1 shows that the constraint binds for a larger set of values of D_t when $\theta = 0.8$ than when $\theta = 1.2$. The left panel of Figure 2 shows, in addition, that noise-trader demand lowers the expected return of an asset with $\theta_n = 0.8$ more than it raises that of an asset with $\theta_n = 1.2$. Hence, uninformed investors should set $\eta < 1$.

The value $\eta = 1$ in Figures 1-3 is optimal in the special case where the probability that uninformed investors give to all values of $\theta \neq 1$ goes to zero. When instead that probability remains positive, uninformed investors set $\eta < 1$, moving the set of allowable positions that managers can take in the risky asset towards lower values. Hence, managers can hold smaller positions in assets in high noise-trader demand, reducing those assets' overvaluation, and weakening the amplification effect that drives up their volatility. Figure 4 shows how the plots in Figures 2 and 3 are modified when $\theta = 1$ has probability 0.6 and each of $\theta = 1 \pm \frac{\Delta\theta}{2}$ for $\Delta\theta \in \{0.1, 0.2, 0.3, 0.4\}$ has probability 0.05. All parameter values except for (η, L), which are chosen optimally, are as in Figures 1-3. We interpret different realizations of θ as different risky assets, and the probability $\pi(\theta)$ of each realization as reflecting the number of assets in supply θ. Thus, there is one asset in supply $\theta_n = 1 \pm \frac{\Delta\theta}{2}$ for $\Delta\theta \in \{0.1, 0.2, 0.3, 0.4\}$, and twelve assets in supply $\theta_n = 1$.

The top left and middle panels of Figure 4 plot expected return and return volatility of two assets as function of the fraction λ of unskilled managers: the asset with $\theta_n = 0.8$ (dashed red line), and the asset with $\theta_{n'} = 1.2$ (solid blue line). Consistent with Figure 2, the difference in expected returns between the two assets increases when λ increases because the constraint tightens. Moreover, the difference in return volatilities is largest for intermediate values of λ, with most of the effect coming from the $\theta_n = 0.8$ asset. The main difference with Figure 2 is that the difference in volatilities is lower: it is approximately 3% when 20% of managers are unskilled, while the maximum difference in Figure 2 is 6%. This reflects the weaker amplification effect.

The top right panel of Figure 4 plots expected return as function of return volatility for $\lambda = 20\%$. The negative relationship derived in Section 3 and plotted in Figure 2 carries through.

The bottom left panel of Figure 4 plots the unconditional averages of the prices of two segments as function of λ: the segment with $(\theta_n, \theta_{n'}) = (0.8, 1.2)$, represented by the thick line, and the segment with $(\theta_{\hat{n}}, \theta_{\hat{n}'}) = (1, 1)$, represented by the thin line. Noise-trader demand is more variable within the former segment. Consistent with Figure 3, that segment trades at a higher price, and its price increases both in absolute terms and relative to the latter segment when λ increases.

The bottom middle panel of Figure 4 plots expected return at the segment level as function of the dispersion in expected returns within the segment for $\lambda = 20\%$ and for the five segments corresponding to $\Delta\theta \in \{0, 0.1, 0.2, 0.3, 0.4\}$. The relationship is weaker than in Figure 4, and turns slightly positive for the first three segments. It remains negative for the last three segments, and is strongly negative for the last two.

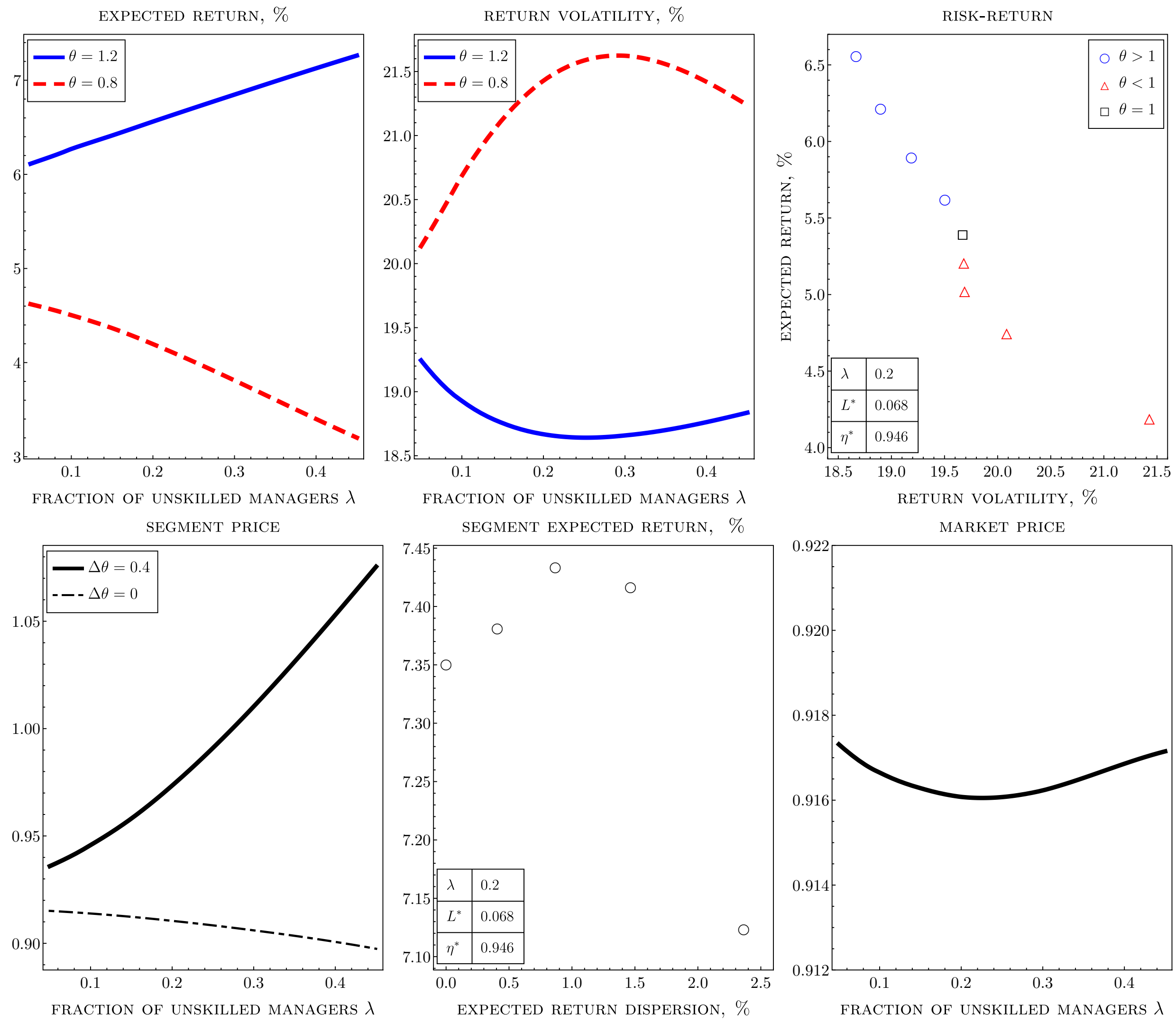
The bottom right panel of Figure 4 plots the unconditional average of the price of a market portfolio in which each asset is in supply of one share. That portfolio is a scaled version of the benchmark portfolio, in which each asset is in supply of η shares. The price is inversely hump-shaped in λ, decreasing slightly for small λ and increasing slightly for larger λ. This result differs from a common finding in previous papers (e.g., Brennan (1993), Kapur and Timmermann (2005), Cuoco and Kaniel (2011), Basak and Pavlova (2013), Buffa and Hodor (2018)) that benchmark-based compensation raises the price of the benchmark portfolio.[11] The price of the benchmark portfolio can drop when managers are more constrained to stay close to that portfolio because of the investors' endogenous response: investors realize that tighter constraints exacerbate price distortions, and reduce their overall investment in the risky assets by lowering η. If we suppress the investors' endogenous response by keeping η independent of λ, then a tighter constraint would raise the price of the benchmark portfolio as in the previous papers. Moreover, the cross-sectional distortions in Figure 4 would become larger and of the same magnitude as those in Figures 2 and 3.

4.3 Effective Capital

Endogenizing the constraint allows us to determine a measure of *effective* capital. According to the market-efficiency view, noise-trader induced distortions should be small because institutions such as mutual funds and pension funds can deploy large pools of capital to trade against them. According to the limits-of-arbitrage view, that capital can be ineffective because agency problems between the

[11] The inverse hump shape in Figure 4 does not change if we multiply the price of the market portfolio by η.

Figure 4: Endogenous Constraint



Expected return and return volatility as functions of λ (top left and middle panels), expected return as function of return volatility for fixed λ (top right panel), segment price and aggregate market price as functions of λ (bottom left and right panels), and segment expected return as function of the dispersion in expected returns within the segment for fixed λ (bottom middle panel). The definitions of these quantities and the drawing conventions are described in the legend of Figure 2 for the top panels, and of Figure 3 for the bottom left and middle panels. There is one asset in supply $\theta_n = 1 \pm \frac{\Delta\theta}{2}$ for $\Delta\theta \in \{0.1, 0.2, 0.3, 0.4\}$, and twelve assets in supply $\theta_n = 1$. The aggregate market price is the unconditional average of the prices of these assets. The values of (η, L) are chosen optimally, and the remaining parameter values are as in Figure 1.

managers in these institutions and the investors who own the capital limit the managers' ability to take risk. The limits-of-arbitrage view emphasizes instead the capital owned by smart-money investors, who are free of agency problems.

An exercise that can inform the debate between the two views is to compute effective capital.

Suppose that a given amount of capital is invested with asset managers. Would price distortions be almost the same as if agency problems between managers and investors did not exist, as per the market-efficiency view? Or would they be larger and comparable to the distortions that would exist if the only capital available were that of smart-money investors, as per the limits-of-arbitrage view? Our model can provide answers to these questions because it incorporates noise traders and price distortions, as well as a constraint on managers that arises from an agency problem with their investors.

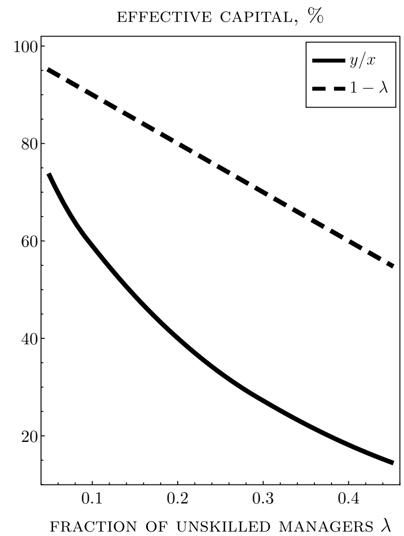
To compute effective capital, we suppose that out of the uninformed investors in measure x, only a subset in measure y can invest but do so with skilled managers to whom they (optimally) impose no constraints. We compute the price distortions under that scenario, and determine the value of y such that distortions are the same as under the equilibrium derived in Section 4.2. That value of y is the "smart-money equivalent" of the x uninformed investors. We refer to y and x as effective and total capital, respectively.

We compute effective capital y in the numerical example of Section 4.2. We measure price distortions by the average difference between the price of assets in high noise-trader demand and the price of assets in low demand. Computing that measure amounts to taking the difference between the price of the asset in supply $\theta_n = 1 - \frac{\Delta\theta}{2}$ and the price of the asset in supply $\theta_n = 1 + \frac{\Delta\theta}{2}$, and averaging over $\Delta\theta \in \{0.1, 0.2, 0.3, 0.4\}$ and D_t.

An increase in the fraction λ of unskilled managers lowers effective capital y through two effects. The direct effect is that that there are fewer skilled managers. The indirect effect is that tighter constraints are imposed on skilled managers. If the direct effect were the only one present, then effective capital would be equal to $(1-\lambda)x$, the measure of investors who employ skilled managers. To compare the direct and indirect effects, we plot in Figure 5 the fraction $\frac{y}{x}$ of effective to total capital (solid line) and the fraction $1-\lambda$ accounted by the direct effect only (dashed line), as functions of λ.

Figure 5 shows that the indirect effect is larger than the direct effect. For example, when $\lambda = 10\%$, effective capital is approximately 60% of total capital, while it should be 90% with only the direct effect present. When $\lambda = 20\%$, effective capital drops to 40%, while under the direct effect it should drop to only 80%. Thus, the constraints imposed on managers reduce significantly the available capital to correct price distortions. At the same time, effective capital remains a non-trivial fraction of total capital.

Figure 5: Effective Capital



Fraction $\frac{y}{x}$ of effective to total capital (solid line) and fraction $1-\lambda$ accounted by the direct effect only (dashed line) as functions of λ. The value of y equates the difference between the price of the asset in supply $\theta_n = 1 - \frac{\Delta\theta}{2}$ and the price of the asset in supply $\theta_n = 1 + \frac{\Delta\theta}{2}$, averaged over $\Delta\theta \in \{0.1, 0.2, 0.3, 0.4\}$ and D_t, to its counterpart in the equilibrium derived in Section 4.2. The values of (η, L) are chosen optimally, and the remaining parameter values are as in Figure 1.

5 Conclusion

We derive equilibrium asset prices when fund managers deviate from benchmark indices to exploit noise-trader induced distortions but fund investors constrain these deviations. Fund managers in our model are not passive, in the sense that they are not constrained to hold benchmark portfolios, nor are they fully active, in the sense that their deviations from these portfolios must lie within bounds. We argue that this view of asset management is more realistic than the conventional active/passive dichotomy, and yields different asset-pricing implications. We develop an approach to endogenize managers' constraints based on investors' uncertainty about managers' skill, and relate the asset-pricing implications to the fraction of unskilled managers.

Our analysis suggests that tracking-error constraints and other portfolio limitations of asset managers can have important effects on portfolio policies and equilibrium asset prices. Empirical research has begun to investigate these effects. For example, Christoffersen and Simutin (2017) find that mutual-fund managers who manage pension-fund assets, and hence face greater pressure to meet benchmarks, hold a larger fraction of their portfolios in high-beta stocks and achieve lower alphas. This is consistent with our result that assets in high noise-trader demand have high betas, and that more constrained managers hold more shares of these assets than less constrained managers. Lines (2016) finds that mutual-fund managers shift their portfolio weights towards the benchmark when volatility rises, putting downward price pressure on overweight stocks and upward pressure on underweight stocks. This is consistent with the amplification effect that we derive.

Extending the empirical investigation by bringing in proxies for noise-trader demand could yield sharper tests of the theoretical mechanisms. Such proxies could include flows into mutual funds, or restricted mandates by institutional investors not to invest in some industry sectors. Empirical studies have documented that high demand according to these proxies is associated with low future returns.[12] Our analysis implies additionally that high demand should be associated with high volatility, and that the trading of managers with tighter constraints should be contributing to this.

A number of extensions are possible on the theoretical front as well. One extension is to make the contracting model between investors and managers dynamic. In a dynamic model, managers could build reputations or could be given incentives based on future termination.[13] Reputations could reduce agency problems. At the same time, they might exacerbate some of the asset-pricing effects that we derive as managers could choose to stay close to their benchmarks to avoid damaging their reputations.

Another extension concerns the normative and policy implications. While each investor in our model seeks to limit the risk taken by his manager, the combined effect of these efforts is to raise the volatility of assets in high noise-trader demand. Would a regulator or a social planner internalize that effect and impose laxer constraints? More generally, how do privately optimal constraints and contracts compare to socially optimal ones? Our setting can help address these questions because it captures the two-way feedback between constraints/contracts and equilibrium asset prices.

[12] Frazzini and Lamont (2008) argue that noise-trader demand ("dumb money" in their terminology) can be proxied by flows into mutual funds, as these predict low long-horizon returns for the stocks bought by the funds. In a similar spirit, Coval and Stafford (2007) find that that stocks sold by mutual funds that experience extreme outflows earn high long-horizon returns, while stocks bought by funds that experience extreme inflows earn low returns. Hong and Kacperczyk (2009) find that stocks in "sin industries" (alcohol, gaming and tobacco) are less held by institutions, presumably because of restricted mandates, and earn higher returns. An alternative proxy for noise-trader demand could be holdings by controlling shareholders, e.g., in family firms.

[13] Papers on fund managers' reputation concerns include Froot, Scharftstein, and Stein (1992), Dasgupta and Prat (2008), Dasgupta, Prat, and Verardo (2011), and Guerrieri and Kondor (2012).

Appendix

A Proofs

Proof of Proposition 3.1. Substituting the affine price function (3.5) into the ODE (3.4), we find

$$D_t + \kappa(\bar{D} - D_t)a_1 - r(a_0 + a_1 D_t) = \rho\theta\sigma^2 D_t a_1^2. \tag{A.1}$$

Equation (A.1) is affine in D_t. Identifying the terms that are linear in D_t yields the equation

$$\rho\theta\sigma^2 a_1^2 + (r + \kappa)a_1 - 1 = 0. \tag{A.2}$$

Equation (A.2) is quadratic in a_1. When $\theta > 0$, the left-hand side is increasing for positive values of a_1, and (A.2) has a unique positive solution, given by (3.7). When $\theta < 0$, the left-hand side is hump-shaped for positive values of a_1, and (A.2) has either two positive solutions, or one positive solution, or no solution. Condition $\theta > -\frac{(r+\kappa)^2}{4\rho\sigma^2}$ in Proposition 3.1 ensures that two positive solutions exist when $\theta < 0$. Equation (3.7) gives the smaller of the two solutions, which is the continuous extension of the unique positive solution when $\theta > 0$. Identifying the constant terms yields the equation

$$\kappa\bar{D}a_1 - ra_0 = 0,$$

whose solution is (3.6).

To show that $S(D_t)$ and $S'(D_t)$ are decreasing and convex in θ, we note that a_1 takes the form

$$\Psi(\theta) \equiv \frac{1}{A + \sqrt{B + C\theta}}$$

for positive constants (A, B, C). The function $\Psi(\theta)$ is decreasing. It is also convex because its derivative

$$\Psi'(\theta) = -\frac{C}{2\sqrt{B + C\theta}} \frac{1}{\left(A + \sqrt{B + C\theta}\right)^2}$$

is increasing. Hence, a_1 is decreasing and convex in θ. These properties extend to a_0 from (3.6), and to $S'(D_t) = a_1$ and $S(D_t) = a_0 + a_1 D_t$. □

Proof of Corollary 3.1. Substituting the price from (3.5) into (3.1), we find that the asset's share return is

$$
\begin{aligned}
dR_t^{sh} &= \left[D_t + \kappa(\bar{D} - D_t)a_1 - r(a_0 + a_1 D_t)\right] dt + \sigma\sqrt{D_t} a_1 dB_t \\
&= \rho\theta\sigma^2 D_t a_1^2 dt + \sigma\sqrt{D_t} a_1 dB_t,
\end{aligned} \tag{A.3}
$$

where the second step follows from (A.1). Substituting the share return from (A.3) and the price from (3.5) into (2.2), we find that the asset's (dollar) return is

$$
\begin{aligned}
dR_t &= \frac{\rho\theta\sigma^2 D_t a_1^2 dt + \sigma\sqrt{D_t} a_1 dB_t}{a_0 + a_1 D_t} \\
&= \frac{\rho\theta\sigma^2 D_t a_1 dt + \sigma\sqrt{D_t} dB_t}{\frac{\kappa}{r}\bar{D} + D_t} \\
&= \frac{\frac{2\rho\theta\sigma^2 D_t dt}{r+\kappa+\sqrt{(r+\kappa)^2+4\rho\theta\sigma^2}} + \sigma\sqrt{D_t} dB_t}{\frac{\kappa}{r}\bar{D} + D_t},
\end{aligned} \tag{A.4}
$$

where the second step follows from (3.6) and the third step follows from (3.7).

The conditional expected return is the drift coefficient in (A.4) times dt,

$$
\mathbb{E}_t(dR_t) = \frac{2\rho\theta\sigma^2 D_t dt}{\left(r + \kappa + \sqrt{(r+\kappa)^2 + 4\rho\theta\sigma^2}\right)\left(\frac{\kappa}{r}\bar{D} + D_t\right)}.
$$

It takes the form $\Phi(\theta)\frac{2\rho\sigma^2 D_t dt}{\frac{\kappa}{r}\bar{D}+D_t}$, where

$$
\Phi(\theta) \equiv \frac{\theta}{A + \sqrt{B + C\theta}}
$$

for positive constants (A, B, C). The function $\Phi(\theta)$ is increasing, and hence the conditional expected return is increasing in θ. (The derivative of $\Phi(\theta)$ has the same sign as

$$
A + \sqrt{B + C\theta} - \frac{C}{2\sqrt{B + C\theta}}\theta = A + \frac{1}{\sqrt{B + C\theta}}\left(B + \frac{C\theta}{2}\right).
$$

This expression is positive for $B + C\theta > 0$, a condition which is required for the term in the square root to be positive.) The unconditional expected return is the unconditional expectation of the conditional expected return,

$$
\mathbb{E}(dR_t) = \mathbb{E}\left(\mathbb{E}_t(dR_t)\right),
$$

because of the law of iterative expectations. Since $\mathbb{E}_t(dR_t)$ is increasing in θ for any given D_t, $\mathbb{E}(dR_t)$ is increasing in θ.

The return's conditional volatility is the diffusion coefficient in (A.4) times $\sqrt{dt}$,

$$\sqrt{\mathbb{V}\text{ar}_t(dR_t)} = \frac{\sigma\sqrt{D_t dt}}{\frac{\kappa}{r}\bar{D} + D_t}. \tag{A.5}$$

It is independent of θ. The return's unconditional variance is the unconditional expectation of the return's conditional variance,

$$\mathbb{V}\text{ar}(dR_t) = \mathbb{E}\left(\mathbb{V}\text{ar}_t(dR_t)\right). \tag{A.6}$$

Since $\mathbb{V}\text{ar}_t(dR_t)$ is independent of θ for any given D_t, $\mathbb{V}\text{ar}(dR_t)$ is independent of θ, and so is the return's unconditional volatility $\sqrt{\mathbb{V}\text{ar}(dR_t)}$. Equation (A.6) is implied by the law of total variance

$$\mathbb{V}\text{ar}(dR_t) = \mathbb{E}\left(\mathbb{V}\text{ar}_t(dR_t)\right) + \mathbb{V}\text{ar}\left(\mathbb{E}_t(dR_t)\right) \tag{A.7}$$

and because in continuous time the second term in the right-hand side of (A.7) is negligible relative to the first: the second term is of order dt^2 while the first is of order dt. $\square$

Proof of Proposition 3.2. Since the ODE (3.8) is identical to (3.4) except that θ is replaced by $\frac{\theta - x\eta}{1-x}$, (3.9) can be derived from (3.7) with the same substitution. The comparisons with the case $L = \infty$ follow because the function $\Psi(\theta)$ defined in the proof of Proposition 3.1 is decreasing. Since $\frac{\theta - x\eta}{1-x} > \theta$ when $\theta > \eta$, (3.7) and (3.9) imply that a_1 is smaller in the case $L = 0$ than in the case $L = \infty$. Conversely, since $\frac{\theta - x\eta}{1-x} < \theta$ when $\theta < \eta$, (3.7) and (3.9) imply that a_1 is larger in the case $L = 0$ than in the case $L = \infty$. These comparisons of a_1 extend to a_0, $S'(D_t) = a_1$ and $S(D_t) = a_0 + a_1 D_t$. $\square$

Proof of Corollary 3.2. The price in the case $L = 0$ can be derived from the price in the case $L = \infty$ by replacing θ by $\frac{\theta - x\eta}{1-x}$. Since the conditional and unconditional volatility in the case $L = \infty$ are independent of θ (Corollary 3.2), they are also independent of θ in the case $L = 0$, and they are equal across the two cases. $\square$

Proof of Theorem 3.1. We prove the theorem through a series of lemmas. Lemma A.1 shows existence of a solution to the ODE system in a compact interval and with initial conditions at the one end of the interval.

Lemma A.1. [Existence in compact interval with conditions at one boundary] *Consider $\epsilon > 0$ and $M > \epsilon$ sufficiently large. A solution $S(D_t)$ to the system of ODEs (3.4) in the* *unconstrained region (3.11), and (3.13) in the constrained region (3.14), with the initial conditions*

$$S'(M) = \frac{2}{r + \kappa + \sqrt{(r+\kappa)^2 + \frac{4\rho(\theta - x\eta)}{1-x}\sigma^2}}, \tag{A.8}$$

$$S(M) = \frac{1}{r}\left((\kappa\bar{D} + rM)S'(M) + \frac{1}{2}\sigma^2 M\Phi + \frac{\rho\, sgn(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M)\right), \tag{A.9}$$

exists, either in the entire interval $[\epsilon, M]$*, or in a maximal interval* $(\hat{\epsilon}, M]$ *with* $\hat{\epsilon} \geq \epsilon$*. In the latter case* $\lim_{D_t \to \hat{\epsilon}} |S'(D_t)| = \infty$*.*

Proof of Lemma A.1. The ODEs (3.4) and (3.13) satisfy the conditions of the Cauchy-Lipschitz theorem for any $D_t > 0$. To show this for the ODE (3.4), we write it as a system of two first-order ODEs:

$$\begin{aligned} S'(D_t) &= T(D_t), \\ T'(D_t) &= \frac{2}{\sigma^2 D_t}\left(\frac{\rho\theta}{1-\lambda x}\sigma^2 D_t T(D_t)^2 - D_t - \kappa(\bar{D} - D_t)T(D_t) + rS(D_t)\right). \end{aligned}$$

The function

$$(D_t, S, T) \longrightarrow \begin{pmatrix} T \\ \frac{2}{\sigma^2 D_t}\left(\frac{\rho\theta}{1-\lambda x}\sigma^2 D_t T^2 - D_t - \kappa(\bar{D} - D_t)T + rS\right) \end{pmatrix}$$

is continuously differentiable for $(D_t, S, T) \in (0,\infty) \times (-\infty,\infty) \times (-\infty,\infty)$. Hence, it is locally Lipschitz in that set, and the Cauchy-Lipschitz theorem implies that for any $(D_t, S, T) \in (0,\infty) \times (-\infty,\infty) \times (-\infty,\infty)$, the ODE (3.4) has a unique solution in a neighborhood of D_t with initial conditions $S(D_t) = S$ and $S'(D_t) = T$. The same argument establishes local existence of a solution to the ODE (3.13).

Consider the solution to the ODE (3.13) with initial conditions (A.8) and (A.9). The value of $S(M)$ in (A.9) is implied from the ODE (3.13) by setting $S''(M) = \Phi$. Indeed, (A.9) is equivalent to

$$S(M) = \frac{1}{r}\left(M + \kappa(\bar{D} - M)S'(M) + \frac{1}{2}\sigma^2 M\Phi - \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 M S'(M)^2 + \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M)\right) \tag{A.10}$$

because the value of $S'(M)$ in (A.8) solves the equation

$$\frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(M)^2 + (r+\kappa)S'(M) - 1 = 0. \tag{A.11}$$

Equation (A.11) is quadratic in $S'(M)$. When $\theta > x\eta$, the left-hand side is increasing for positive values of $S'(M)$, and (A.11) has a unique positive solution, given by (A.8). When $\theta < x\eta$, the left-hand side is hump-shaped for positive values of $S'(M)$, and (A.11) has either two positive solutions, or one positive solution, or no solution. Condition $\theta > x\eta - \frac{(1-x)(r+\kappa)^2}{4\rho\sigma^2}$ in Theorem 3.1 ensures that two positive solutions exist when $\theta < x\eta$. Equation (A.8) gives the smaller of the two solutions, which is the continuous extension of the unique positive solution when $\theta > x\eta$.

Since $S'(M)$ is independent of M, (3.14) is met for M sufficiently large. Continuity then implies that the solution to the ODE (3.13) with initial conditions (A.8) and (A.9) lies in the constrained region (3.14) in a neighborhood to the left of M. We extend the solution maximally to the left of M, up to a point m_1 where either the solution explodes ($\lim_{D_t \to m_1} |S'(D_t)| = \infty$) or condition (3.14) that defines the constrained region is violated in a neighborhood to the left of m_1. In the second case, we extend the solution to the left of m_1 by using the ODE (3.4) instead of (3.13). If the first derivative of $\sqrt{D_t}S'(D_t)$ at m_1 is non-zero, then it has to be positive because (3.14) is violated to the left of m_1, and the extended solution lies in the unconstrained region (3.11) in a neighborhood to the left of m_1, by continuity. (Extending the solution to the left of m_1 by using the ODE (3.4) instead of (3.13) yields the same first derivative of $\sqrt{D_t}S'(D_t)$, i.e., the first derivatives of $\sqrt{D_t}S'(D_t)$ from the left, using (3.4), and the right, using (3.13), coincide. The first derivatives of $S(D_t)$ from the left and the right coincide because the first derivative from the right is used as initial condition when extending the solution to the left. The second derivatives of $S(D_t)$ from the left and the right coincide because the first derivatives coincide and (3.11) holds with equality at m_1. The result, used next in the proof, that higher-order derivatives of $\sqrt{D_t}S'(D_t)$ from the left and the right coincide if all lower-order derivatives are zero uses a similar argument and differentiation of (3.4) and (3.13).) If the first derivative of $\sqrt{D_t}S'(D_t)$ at m_1 is zero, then the second derivative must also be zero because otherwise (3.14) would not be violated to the left of m_1. If the third derivative of $\sqrt{D_t}S'(D_t)$ at m_1 is non-zero, then it has to be positive because (3.14) is violated to the left of m_1, and the extended solution lies in the unconstrained region (3.11) in a neighborhood to the left of m_1, by continuity. Proceeding in this manner for higher-order derivatives, we conclude that the extended solution (using the ODE (3.4) instead of (3.13) to the left of m_1) may not lie in the unconstrained region (3.11) in a neighborhood to the left of m_1 only if all n-th order derivatives of $\sqrt{D_t}S'(D_t)$ at m_1, for $n \geq 1$, are zero. Writing, however, the ODE (3.13) in terms of the function $U(D_t) \equiv \sqrt{D_t}S'(D_t)$ taking the $(n+1)$-th order derivative of the resulting equation at m_1, and using $U(m_1) > 0$ and $\frac{d^{n+1}}{dD_t^{n+1}} [U(D_t)]_{D_t=m_1} = 0$ for all $n \geq 0$, we find

$$\frac{d^{n+1}}{dD_t^{n+1}} \left[D_t + \kappa(\bar{D} - D_t)\frac{U(D_t)}{\sqrt{D_t}} + \frac{1}{2}\sigma^2 \sqrt{D_t} \left(U'(D_t) - \frac{1}{2D_t} U(D_t) \right) - rS(M) \right.$$
$$\left. -r \int_M^{D_t} \frac{U(D_t')}{\sqrt{D_t'}} dD_t' \right]_{D_t=m_1} = \frac{d^{n+1}}{dD_t^{n+1}} \left[\frac{\rho(\theta - x\eta)}{1-x} \sigma^2 U(D_t)^2 - \frac{\rho \operatorname{sgn}(\theta - \eta) x L}{1-x} \sigma U(D_t) \right]_{D_t=m_1}$$
$$\Rightarrow \frac{d^{n+1}}{dD_t^{n+1}} \left[\kappa \frac{\bar{D} - D_t}{\sqrt{D_t}} - \frac{1}{4}\sigma^2 \frac{1}{\sqrt{D_t}} \right]_{D_t=m_1} - r \frac{d^n}{dD_t^n} \left[\frac{1}{\sqrt{D_t}} \right]_{D_t=m_1} = 0$$

for all $n \geq 0$, a contradiction. Hence, the extended solution lies in the unconstrained region (3.11) in a neighborhood to the left of m_1. We extend that solution maximally to the left of m_1, up to a point m_2 where either the solution explodes ($\lim_{D_t \to m_2} |S'(D_t)| = \infty$) or where condition (3.11) is violated in a neighborhood to the left of m_2. In the second case, we extend the solution to the left of m_1 by using the ODE (3.13) instead of (3.4). Repeating this process yields a solution to the system of ODEs (3.4) in the unconstrained region (3.11), and (3.13) in the constrained region (3.14), with initial conditions (A.8) and (A.9), which either is defined in $[\epsilon, M]$ or explodes at an $\hat{\epsilon} \geq \epsilon$. □

Lemma A.2 shows that the solution derived in Lemma A.1 is either increasing in D_t or is decreasing and then increasing.

Lemma A.2. [Monotonicity] *For the solution derived in Lemma A.1, either $S'(D_t) > 0$ for all D_t, or there exists $m < M$ such that $S'(D_t) > 0$ for all $D_t \in (m, M]$, $S'(D_t) < 0$ for all $D_t < m$, and $S(m) > 0$.*

Proof of Lemma A.2. Since $S'(M) > 0$, $S'(D_t) > 0$ for D_t smaller than and close to M. Suppose that there exists $D_t < M$ such that $S'(D_t) \leq 0$, and consider the supremum m within that set. The definition of m implies $S'(D_t) > 0$ for all D_t in the non-empty set (m, M), $S'(m) = 0$, and $S''(m) \geq 0$. If $S''(m) = 0$, then differentiation of (3.4) and (3.13) at m yields $S'''(m) < 0$, which contradicts $S'(D_t) > S'(m) = 0$ for $D_t > m$. Hence, $S''(m) > 0$, which in turn implies $S'(D_t) < 0$ for D_t smaller than and close to M

Suppose next, by contradiction, that there exists $D_t < m$ such that $S'(D_t) \geq 0$, and consider the supremum m_1 within that set. The definition of m_1 implies that $S'(D_t) < 0$ for all D_t in the non-empty set (m_1, m), $S'(m_1) = 0$, and $S''(m_1) \leq 0$.

Substituting $S'(m) = 0$ and $S''(m) > 0$ in (3.4) and (3.13), we find that in both cases

$$m - rS(m) < 0. \tag{A.12}$$

Likewise, substituting $S'(m_1) = 0$ and $S''(m_1) \leq 0$ in (3.4) and (3.13), we find

$$m_1 - rS(m_1) \geq 0. \tag{A.13}$$

Equations (A.12) and (A.13) imply

$$S(m) - S(m_1) > \frac{m - m_1}{r} > 0,$$

which contradicts $S'(D_t) < 0$ for all $D_t \in (m_1, m)$. Hence, either $S'(D_t) > 0$ for all D_t, or there exists $m < M$ such that $S'(D_t) > 0$ for all $D_t \in (m, M]$ and $S'(D_t) < 0$ for all $D_t < m$. In the latter case, (A.12) implies $S(m) > \frac{m}{r} > 0$. $\square$

Lemma A.3 shows a monotonicity property of the solution with respect to the initial conditions. If a solution $S_1(D_t)$ lies below another solution $S_2(D_t)$ at M, and their first derivatives are equal at M, then $S_1(D_t)$ lies below $S_2(D_t)$ for all $D_t < M$, while the comparison reverses for the first derivatives.

Lemma A.3. [**Monotonicity over initial conditions**] *Consider two solutions $S_1(D_t)$ and $S_2(D_t)$ derived in Lemma A.1 for Φ_1 and $\Phi_2 > \Phi_1$, respectively. For all $D_t < M$, $S_1(D_t) < S_2(D_t)$ and $S_1'(D_t) > S_2'(D_t)$.*

Proof of Lemma A.3. Equation (A.8) implies $S_1'(M) = S_2'(M)$. Equations (A.10) and $\Phi_1 < \Phi_2$ imply $S_1(M) < S_2(M)$ and $S_1''(M) < S_2''(M)$. Combining the latter inequality with $S_1'(M) = S_2'(M)$, we find $S_1'(D_t) > S_2'(D_t)$ for D_t smaller than and close to M. Moreover, by continuity, $S_1(D_t) < S_2(D_t)$ for D_t smaller than and close to M. Suppose, by contradiction, that there exists $D_t < M$ such that $S_1(D_t) \geq S_2(D_t)$ or $S_1'(D_t) \leq S_2'(D_t)$, and consider the supremum m within that set. The definition of m implies $S_1(D_t) < S_2(D_t)$ and $S_1'(D_t) > S_2'(D_t)$ for all D_t in the non-empty set (m, M), and $S_1(m) = S_2(m)$ or $S_1'(m) = S_2'(m)$.

Since $S_1(M) < S_2(M)$ and $S_1'(D_t) > S_2'(D_t)$ for all $D_t \in (m, M)$, $S_1(m) < S_2(m)$. Hence, $S_1'(m) = S_2'(m)$. Equations (3.4) and (3.13) both imply, however, that since $S_1(m) < S_2(m)$, $S_1''(m) < S_2''(m)$. Hence, $S_1'(D_t) < S_2'(D_t)$ for D_t close to and larger than m, a contradiction. $\square$

Lemma A.4 derives properties of the solution for $\Phi = 0$. For this and subsequent results, we use the function $Z(D_t)$ defined by

$$Z(D_t) \equiv (\kappa \bar{D} + rD_t)S'(D_t) - rS(D_t).$$

Lemma A.4. [Solution for $\Phi = 0$] *The solution $S(D_t)$ derived in Lemma A.1 has the following properties for $\Phi = 0$:*

- *When $\theta > \eta$, the solution satisfies $Z(\epsilon) < 0$ if it can be defined in $[\epsilon, M]$, and satisfies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$ and $\lim_{D_t \to \hat{\epsilon}} S(D_t) > 0$ if it explodes at $\hat{\epsilon} \geq \epsilon$.*
- *When $\theta < \eta$, the solution can be defined in $[\epsilon, M]$, and satisfies $Z(\epsilon) > 0$.*

Proof of Lemma A.4. We start with the case $\theta > \eta$. Suppose first that there exists $D_t < M$ such that $S'(D_t) \leq 0$. Lemma A.2 implies that there exists a unique $m < M$ such that $S'(D_t) > 0$ for all $D_t \in (m, M]$, $S'(D_t) < 0$ for all $D_t < m$, and $S(m) > 0$. Hence, if the solution can be defined in $[\epsilon, M]$, it satisfies

$$(\kappa\bar{D} + r\epsilon)S'(\epsilon) \leq 0 < rS(m) \leq rS(\epsilon),$$

which implies $Z(\epsilon) < 0$. If instead the solution explodes at $\hat{\epsilon} \geq \epsilon$, it satisfies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$ and $\lim_{D_t \to \hat{\epsilon}} S(D_t) > S(m) > 0$.

Suppose next that $S'(D_t) > 0$ for all $D_t \leq M$. We will show that the solution is convex, can be defined in $[\epsilon, M]$, and satisfies $Z(\epsilon) < 0$. We first show that $S'''(M) < 0$. We write the ODE (3.13) as

$$\frac{1}{2}\sigma^2 S''(D_t) = \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(D_t)^2 - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{\sqrt{D_t}}S'(D_t) - 1 + \frac{rS(D_t) - \kappa\bar{D}S'(D_t)}{D_t} + \kappa S'(D_t). \tag{A.14}$$

Differentiating both sides, we find

$$\begin{aligned}\frac{1}{2}\sigma^2 S'''(D_t) =& 2\frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(D_t)S''(D_t) - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\left(\frac{1}{\sqrt{D_t}}S''(D_t) - \frac{1}{2D_t^{\frac{3}{2}}}S'(D_t)\right)\\ &+ \frac{rS'(D_t) - \kappa\bar{D}S''(D_t)}{D_t} - \frac{rS(D_t) - \kappa\bar{D}S'(D_t)}{D_t^2} + \kappa S''(D_t).\end{aligned} \tag{A.15}$$

Setting $D_t = M$ in (A.15) and using $S''(M) = \Phi = 0$, we find

$$\begin{aligned}\frac{1}{2}\sigma^2 S'''(M) &= \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2M^{\frac{3}{2}}}S'(M) + \frac{rS'(M)}{M} - \frac{rS(M) - \kappa\bar{D}S'(M)}{M^2}\\ &= -\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2M^{\frac{3}{2}}}S'(M) < 0,\end{aligned} \tag{A.16}$$

where the second step follows by substituting $S(M)$ from (A.9) and using again $\Phi = 0$.

Since $S'''(M) < 0$ and $S''(M) = 0$, $S''(D_t) > 0$ for D_t smaller than and close to M. Suppose, by contradiction, that there exists $D_t < M$ such that $S''(D_t) \leq 0$, and consider the supremum m within that set. The definition of m implies that $S''(D_t) > 0$ for all D_t in the non-empty set (m, M), $S''(m) = 0$, and $S'''(m) \geq 0$.

Suppose that m lies in the constrained region. Setting $D_t = m$ in (A.15), and using $S''(m) = 0$ and $S'''(m) \geq 0$, we find

$$\begin{aligned}
&\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2m^{\frac{3}{2}}}S'(m) + \frac{rS'(m)}{m} - \frac{rS(m)-\kappa\bar{D}S'(m)}{m^2} \geq 0 \\
\Leftrightarrow &-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2m^{\frac{3}{2}}}S'(m) + \frac{1}{m}\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) \geq 0,
\end{aligned} \tag{A.17}$$

where the second step follows by substituting $S(m)$ from (3.13) and using again $S''(m) = 0$. The contradiction follows because both terms in the left-hand side of (A.17) are negative. The first term is negative because $S'(m) > 0$. The second term is negative because (i) $S''(D_t) > 0$ for all $D_t \in (m, M)$ implies $S'(m) < S'(M)$, and (ii) the latter inequality together with $S'(m) > 0$ imply that the left-hand side of (A.11) becomes negative when $S'(M)$ is replaced by $S'(m)$.

Suppose next that m lies in the unconstrained region. The ODE (3.4) yields the following counterpart of (A.15):

$$\frac{1}{2}\sigma^2 S'''(D_t) = 2\rho\theta\sigma^2 S'(D_t)S''(D_t) + \frac{rS'(D_t)-\kappa\bar{D}S''(D_t)}{D_t} - \frac{rS(D_t)-\kappa\bar{D}S'(D_t)}{D_t^2} + \kappa S''(D_t). \tag{A.18}$$

Setting $D_t = m$ in (A.18), and using $S''(m) = 0$, $S'''(m) \geq 0$, and (3.4), we find the following counterpart of (A.17):

$$\frac{1}{m}\left(\rho\theta\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) \geq 0. \tag{A.19}$$

The contradiction follows because (i) $S''(D_t) > 0$ for all $D_t \in (m, M)$ implies $S'(m) < S'(M)$, (ii) the latter inequality together with $S'(m) > 0$ imply that the left-hand side of (A.11) becomes negative when $S'(M)$ is replaced by $S'(m)$, and (iii) the left-hand side of (A.11) being negative and $\theta > \eta$ imply that the left-hand side of (A.19) is negative. Since $S''(D_t) > 0$ for all $D_t < M$, $S(D_t)$ is convex.

If the solution explodes at $\hat{\epsilon} \geq \epsilon$, then convexity implies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$, contradicting

$S'(D_t) > 0$ for all D_t. Hence, the solution can be defined in $[\epsilon, M]$. Moreover, convexity implies

$$\begin{aligned} rS(\epsilon) &\geq rS(M) + r(\epsilon - M)S'(M) \\ &= (\kappa\bar{D} + r\epsilon)S'(M) + \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M), \end{aligned} \tag{A.20}$$

where the second step follows by substituting $S(M)$ from (A.9) and using $\Phi = 0$. Equation (A.20) implies $Z(\epsilon) < 0$ because $S'(M) > 0$ and $S'(M) > S'(\epsilon)$.

We next consider the case $\theta < \eta$. We will show that the solution is concave, can be defined in $[\epsilon, M]$, and satisfies $Z(\epsilon) > 0$. Equation (A.16) implies $S'''(M) > 0$. Since $S'''(M) > 0$ and $S''(M) = 0$, $S''(D_t) < 0$ for D_t smaller than and close to M. Suppose, by contradiction, that there exists $D_t < M$ such that $S''(D_t) \geq 0$, and consider the supremum m within that set. The definition of m implies that $S''(D_t) < 0$ for all D_t in the non-empty set (m, M), $S''(m) = 0$, and $S'''(m) \leq 0$.

Suppose that m lies in the unconstrained region. Since $S''(m) = 0$ and $S'''(m) \leq 0$, (A.19) holds as an inequality in the opposite direction, i.e.,

$$\frac{1}{m}\left(\rho\theta\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) \leq 0. \tag{A.21}$$

When $\eta > \theta > x\eta$, (A.21) yields a contradiction because (i) $S''(D_t) < 0$ for all $D_t \in (m, M)$ implies $S'(m) > S'(M)$, (ii) the latter inequality implies that the left-hand side of (A.11) becomes positive when $S'(M)$ is replaced by $S'(m)$, and (iii) the left-hand side of (A.11) being positive and $\theta < \eta$ imply that the left-hand side of (A.21) is positive. When, however, $x\eta > \theta$, (A.21) does not yield a contradiction because the left-hand side of (A.11) is hump-shaped for positive values of $S'(M)$, rather than increasing. It increases until the mid-point between the two positive roots, and then decreases to $-\infty$.

To derive a contradiction when $x\eta > \theta$, we examine the behavior of $Z(D_t)$ in (m, M). Since

$$Z'(D_t) = (\kappa\bar{D} + rD_t)S''(D_t) < 0 \tag{A.22}$$

for all $D_t \in (m, M)$, $Z(D_t)$ is decreasing in (m, M). Moreover, (A.9) and $\Phi = 0$ imply

$$Z(M) = (\kappa\bar{D} + rM)S'(M) - rS(M) = -\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M), \tag{A.23}$$

and (3.4) and $S''(m) = 0$ imply

$$Z(m) = (\kappa\bar{D} + rm)S'(m) - rS(m) = m\left(\rho\theta\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right). \tag{A.24}$$

Since $Z(D_t)$ is decreasing,

$$\begin{aligned}
&Z(m) > Z(M)\\
&\Leftrightarrow m\left(\rho\theta\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) > -\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M) > 0,
\end{aligned} \tag{A.25}$$

where the second step follows from (A.23) and (A.24). Equation (A.25) contradicts (A.21).

Suppose next that m lies in the constrained region. Since $S''(m) = 0$ and $S'''(m) \leq 0$, (A.17) holds as an inequality in the opposite direction, i.e.,

$$-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2m^{\frac{3}{2}}}S'(m) + \frac{1}{m}\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) \leq 0. \tag{A.26}$$

When $\eta > \theta > x\eta$, (A.26) yields a contradiction because both terms in the left-hand side are positive. The first term is positive because $S'(m) > 0$. The second term is positive because (i) $S''(D_t) < 0$ for all $D_t \in (m, M)$ implies $S'(m) > S'(M)$, and (ii) the latter inequality implies that the left-hand side of (A.11) becomes positive when $S'(M)$ is replaced by $S'(m)$. To derive a contradiction when $x\eta > \theta$, we examine the behavior of $Z(D_t)$ in (m, M). Equations (3.13) and $S''(m) = 0$ imply

$$\begin{aligned}
Z(m) &= (\kappa\bar{D} + rm)S'(m) - rS(m)\\
&= m\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{m}S'(m).
\end{aligned} \tag{A.27}$$

Equation (A.22) implies

$$\begin{aligned}
&Z(m) = Z(M) - \int_m^M (\kappa\bar{D} + rD_t)S''(D_t)dD_t\\
&\Rightarrow Z(m) \geq Z(M) - \int_m^M rmS''(D_t)dD_t\\
&\Leftrightarrow Z(m) > Z(M) + rm[S'(m) - S'(M)]\\
&\Leftrightarrow m\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1\right) - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{m}S'(m)\\
&> -\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M) + rm[S'(m) - S'(M)],
\end{aligned} \tag{A.28}$$

where the last step follows from (A.23) and (A.27). Combining (A.26) and (A.28), we find

$$\begin{aligned}
&-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{\sqrt{m}}{2}S'(m) > -\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M) + rm[S'(m) - S'(M)]\\
&\Leftrightarrow \frac{\rho xL}{1-x}\sigma\frac{\sqrt{m}}{2}S'(m) - rm[S'(m) - S'(M)] > \frac{\rho xL}{1-x}\sigma\sqrt{M}S'(M).
\end{aligned} \tag{A.29}$$

The left-hand side of (A.29) is linear in $S'(m)$. Since $S''(D_t) < 0$ for all $D_t \in (m, M)$, $S'(m)$ is bounded below by $S'(M)$. To derive an upper bound for $S'(m)$, we note that since $S'(m) > S'(M)$, (A.28) implies

$$\frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m)^2 + (r+\kappa)S'(m) - 1 + \frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{m}}S'(m) > 0.$$

Hence, $S'(m)$ is smaller than the larger positive root of the quadratic equation

$$\frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m)^2 + \left(r + \kappa + \frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{m}}\right)S'(m) - 1 = 0,$$

which is

$$-\frac{\left(r+\kappa+\frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{m}}\right) + \sqrt{\left(r+\kappa+\frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{m}}\right)^2 + 4\frac{\rho(\theta-x\eta)\sigma^2}{1-x}}}{2\frac{\rho(\theta-x\eta)\sigma^2}{1-x}}.$$

This root is, in turn, smaller than

$$-\frac{\left(r+\kappa+\frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{m}}\right) + \left(r+\kappa+\frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{m}}\right)\sqrt{1 + \frac{4\frac{\rho(\theta-x\eta)\sigma^2}{1-x}}{(r+\kappa)^2}}}{2\frac{\rho(\theta-x\eta)\sigma^2}{1-x}} = S^* + \frac{B}{\sqrt{m}},$$

where S^* is the larger positive root of (A.11) and

$$B \equiv -\frac{\frac{\rho x L}{1-x}\sigma + \frac{\rho x L}{1-x}\sqrt{1 + \frac{4\frac{\rho(\theta-x\eta)\sigma^2}{1-x}}{(r+\kappa)^2}}}{2\frac{\rho(\theta-x\eta)\sigma^2}{1-x}} > 0.$$

When $S'(m)$ in (A.29) is set to $S'(M)$, the left-hand side is smaller than the right-hand side. When $S'(m)$ in (A.29) is set to the upper bound $S^* + \frac{B}{\sqrt{m}}$, the left-hand side is a quadratic function of $\sqrt{m}$, with the coefficient of $(\sqrt{m})^2 = m$ being $-r[S^* - S'(M)] < 0$. It is, therefore, bounded above, and smaller than the right-hand side for sufficiently large M. Hence, (A.29) does not hold, a contradiction. Since $S''(D_t) < 0$ for all $D_t < M$, $S(D_t)$ is concave.

If the solution explodes at $\hat{\epsilon} \geq \epsilon$, then concavity implies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = \infty$. The right-hand side of (3.4) and (3.13) is of order $S'(D_t)^2$ for D_t close to $\hat{\epsilon}$. The left-hand side, however, does not exceed

$$\begin{aligned}
&D_t + \kappa(\bar{D} - D_t)S'(D_t) - rS(D_t) \\
&\leq D_t + \kappa(\bar{D} - D_t)S'(D_t) - rS(M) - r(D_t - M)S'(D_t),
\end{aligned}$$

where both the first and the second steps follow from concavity. Hence, the left-hand side is bounded by a term of order $S'(D_t)$, a contradiction. Therefore, the solution does not explode and can be defined in $[\epsilon, M]$. Equation $Z(\epsilon) > 0$ holds because $Z(D_t)$ is decreasing and $Z(M) > 0$ from (A.23). $\square$

Lemma A.5 derives properties of the solution for $|\Phi|$ large.

Lemma A.5. [Solution for large $|\Phi|$] *The solution $S(D_t)$ derived in Lemma A.1 has the following properties:*

- *When $\theta > \eta$ and Φ is negative and large, the solution can be defined in $[\epsilon, M]$, and satisfies $Z(\epsilon) > 0$.*
- *When $\theta < \eta$ and Φ is positive and large, the solution satisfies $Z(\epsilon) < 0$ if it can be defined in $[\epsilon, M]$, and satisfies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$ and $\lim_{D_t \to \hat{\epsilon}} S(D_t) > 0$ if it explodes at $\hat{\epsilon} \geq \epsilon$.*

Proof of Lemma A.5. We start with the case $\theta > \eta$. Suppose that Φ is negative and sufficiently large so that $S(M)$ defined by (A.9) is negative. We will show that $S'(D_t) > 0$ and $S''(D_t) < 0$ for all D_t. Both inequalities hold by continuity for D_t smaller than and close to M. Suppose, by contradiction, that there exists $D_t < M$ such that $S'(D_t) \leq 0$ or $S''(D_t) \geq 0$, and consider the supremum m within that set. The definition of m implies $S'(D_t) > 0$ and $S''(D_t) < 0$ for all D_t in the non-empty set (m, M), and $S'(m) = 0$ or $S''(m) = 0$.

Since $S'(M) > 0$ and $S''(D_t) < 0$ for all $D_t \in (m, M)$, $S'(m) > 0$. Hence, $S''(m) = 0$. Since, in addition, $S(M) < 0$ and $S'(D_t) > 0$ for all $D_t \in (m, M)$, $S(m) < 0$. Setting $D_t = m$ in (A.15) and (A.18), and using $S(m) < 0$, $S'(m) > 0$ and $S''(m) = 0$, we find $S'''(m) > 0$. Hence, $S''(D_t) > 0$ for D_t close to and larger than m, a contradiction. Therefore, $S'(D_t) > 0$ and $S''(D_t) < 0$ for all D_t.

Since the solution is concave, we can use the same argument as in the proof of Lemma A.4 in the case $\theta < 0$, to show that the solution does not explode at $\hat{\epsilon} \geq \epsilon$. Hence, the solution can be defined in $[\epsilon, M]$. It satisfies $Z(\epsilon) > 0$ because $S(\epsilon) < 0$ and $S'(\epsilon) > 0$.

We next consider the case $\theta < \eta$. We will show, by contradiction, that there exists $D_t < M$ such that $S'(D_t) \leq 0$. Existence of such a D_t will imply, from Lemma A.2, existence of a unique $m < M$ such that $S'(D_t) > 0$ for all $D_t \in (m, M]$, $S'(D_t) < 0$ for all $D_t < m$, and $S(m) > 0$. Hence, if the solution can be defined in $[\epsilon, M]$, it satisfies

$$(\kappa\bar{D} + r\epsilon)S'(\epsilon) \leq 0 < rS(m) \leq rS(\epsilon),$$

which implies $Z(\epsilon) < 0$. If instead the solution explodes at $\hat{\epsilon} \geq \epsilon$, it satisfies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$ and $\lim_{D_t \to \hat{\epsilon}} S(D_t) > S(m) > 0$.

To derive the contradiction, we assume that $S'(D_t) > 0$ for all $D_t \leq M$, and will show that $S''(D_t)$ is bounded below by $\frac{\Phi}{2}$. Continuity yields the bound $S''(D_t) \geq \frac{\Phi}{2}$ for D_t smaller than and close to M because $S'(M) = \Phi$. Suppose, by contradiction, that there exists D_t such that $S''(D_t) < \frac{\Phi}{2}$, and consider the supremum within that set. The definition of m implies $S''(m) > \frac{\Phi}{2}$ for all D_t in the non-empty set (m, M), and $S''(m) = \frac{\Phi}{2}$.

If m lies in the constrained region, (3.13) implies

$$\begin{aligned}
\frac{1}{2}\sigma^2 S''(m) &= \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m)^2 - \frac{\rho \operatorname{sgn}(\theta-\eta) x L}{1-x}\sigma \frac{1}{\sqrt{m}} S'(m) - 1 + \frac{rS(m) - \kappa(\bar{D} - m)S'(m)}{m} \\
&\geq \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m)^2 - 1 + \frac{rS(M) + rS'(M)(m-M) - \kappa(\bar{D} - m)S'(m)}{m} \\
&= \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m)^2 - 1 \\
&\quad + \frac{\kappa\bar{D}(S'(M) - S'(m)) + \frac{1}{2}\sigma^2 M\Phi + \frac{\rho \operatorname{sgn}(\theta-\eta) x L}{1-x}\sigma\sqrt{M}S'(M) + rmS'(M) + \kappa m S'(m)}{m} \\
&> \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(M)^2 - 1 + \frac{\frac{1}{2}\sigma^2 M\Phi - \frac{\rho x L}{1-x}\sigma\sqrt{M}S'(M)}{m},
\end{aligned} \tag{A.30}$$

where the second step follows from $S'(m) > 0$ and because convexity implies

$$S(m) \geq S(M) + S'(M)(m - M),$$

the third step follows by substituting $S(M)$ from (A.9), and the fourth step follows because $S'(M) > S'(m) > 0$. Since for sufficiently large Φ,

$$\frac{1}{2}\sigma^2 M\Phi - \frac{\rho x L}{1-x}\sigma\sqrt{M}S'(M) > 0,$$

the right-hand side of (A.30) is bounded below by

$$\frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(M)^2 - 1 + \frac{1}{2}\sigma^2\Phi - \frac{\rho x L}{1-x}\sigma\frac{1}{\sqrt{M}}S'(M),$$

which, in turn, is bounded below by $\frac{1}{4}\sigma^2\Phi$ for sufficiently large Φ. Hence, (A.30) implies that $S''(m)$ exceeds $\frac{\Phi}{2}$, a contradiction. If m lies in the unconstrained region, we can follow the same steps to derive a counterpart of (A.30) using (3.4), and then derive a contradiction. Hence, $S''(D_t) \geq \frac{\Phi}{2}$ for all $D_t \leq M$.

If the solution explodes at $\hat{\epsilon} \geq \epsilon$, then convexity implies $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$. This is ruled out, however, by $S'(D_t) > 0$ for all $D_t \leq M$. Hence, the solution can be defined in $[\epsilon, M]$.

Since, however, $S''(D_t)$ is bounded below by $\frac{\Phi}{2}$, and $S'(M)$ is independent of Φ, $S'(\epsilon)$ is negative for sufficiently large Φ. This contradicts our assumption that $S'(D_t) > 0$ for all $D_t \leq M$, and establishes that there exists $D_t < M$ such that $S'(D_t) \leq 0$. $\square$

Taken together, Lemmas A.4 and A.5 show that for two extreme values of Φ ($\Phi = 0$ and $|\Phi|$ large) the solution lies on two different "sides" of the equation $Z(\epsilon) = 0$, which we use as boundary condition at ϵ. Lemma A.6 uses these results and a continuity argument to show that there exists Φ such that $Z(\epsilon) = 0$ holds. It also uses the monotonicity property of the solution shown in Lemma A.3 to establish that this Φ is unique.

Lemma A.6. [Existence in compact interval with conditions at both boundaries] *Consider an interval $[\epsilon, M]$, with ϵ sufficiently small and M sufficiently large. A solution $S(D_t)$ to the system of ODEs (3.4) in the unconstrained region (3.11), and (3.13) in the constrained region (3.14), with the boundary conditions (A.8) and $Z(\epsilon) = 0$ exists in $[\epsilon, M]$ and is unique. Moreover, $S''(M) < 0$ when $\theta > \eta$, and $S''(M) > 0$ when $\theta < \eta$.*

Proof of Lemma A.6. We denote by $Z_\Phi(\epsilon)$ the value of $Z(\epsilon)$ for the solution $S(D_t)$ derived in Lemma A.1. If $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = -\infty$ for $\hat{\epsilon} \geq \epsilon$, in which case $\lim_{D_t \to \hat{\epsilon}} S(D_t) > 0$, we set $Z_\Phi(\epsilon) = -\infty$. If $\lim_{D_t \to \hat{\epsilon}} S'(D_t) = \infty$ for $\hat{\epsilon} \geq \epsilon$, in which case $\lim_{D_t \to \hat{\epsilon}} S(D_t)$ is finite or $-\infty$, we set $Z_\Phi(\epsilon) = \infty$.

Lemma A.3 implies that for $\Phi_1 < \Phi_2$, $Z_{\Phi_1}(\epsilon) > Z_{\Phi_2}(\epsilon)$ if $Z_{\Phi_1}(\epsilon)$ and Z_{Φ_2} are finite, $Z_{\Phi_2}(\epsilon) = -\infty$ if $Z_{\Phi_1}(\epsilon) = -\infty$, and $Z_{\Phi_1}(\epsilon) = \infty$ if $Z_{\Phi_2}(\epsilon) = \infty$. Hence, $Z_\Phi(\epsilon)$ is equal to ∞ in an interval $(-\infty, \underline{\Phi}]$, is finite and decreasing in an interval $(\underline{\Phi}, \bar{\Phi})$, and is equal to $-\infty$ in the remaining interval $[\bar{\Phi}, \infty)$. Continuity of the solution with respect to the initial conditions implies that $Z_\Phi(\epsilon)$ is continuous in Φ in $(\underline{\Phi}, \bar{\Phi})$. Moreover, if $\underline{\Phi}$ is finite, then $\lim_{\substack{\Phi \to \underline{\Phi} \\ \Phi > \underline{\Phi}}} Z_\Phi(\epsilon) = \infty$, and if $\bar{\Phi}$ is finite, then $\lim_{\substack{\Phi \to \bar{\Phi} \\ \Phi < \bar{\Phi}}} Z_\Phi(\epsilon) = -\infty$.

When $\theta > \eta$, Lemma A.4 implies that $Z_0(\epsilon)$ is negative and possibly equal to $-\infty$, and Lemma A.5 implies that $Z_\Phi(\epsilon)$ is positive and finite for Φ negative and large. Hence, $\underline{\Phi} = -\infty$. If $\bar{\Phi} > 0$, then continuity and monotonicity of $Z_\Phi(\epsilon)$ in $(-\infty, 0]$, $\lim_{\Phi \to -\infty} Z_\Phi(\epsilon) > 0$, and $Z_0(\epsilon) < 0$ imply that there exists a unique $\Phi \in (-\infty, 0)$ such that $Z_\Phi(\epsilon) = 0$. If $\bar{\Phi} \leq 0$, then continuity and monotonicity of $Z_\Phi(\epsilon)$ in $(-\infty, \bar{\Phi})$, $\lim_{\Phi \to -\infty} Z_\Phi(\epsilon) > 0$, and $\lim_{\substack{\Phi \to \bar{\Phi} \\ \Phi < \bar{\Phi}}} Z_\Phi(\epsilon) = -\infty$ imply that there exists a unique $\Phi \in (-\infty, \bar{\Phi})$ such that $Z_\Phi(\epsilon) = 0$. In both cases, there exists a unique $\Phi \in (-\infty, 0)$ such that $Z_\Phi(\epsilon) = 0$. The solution $S(D_t)$ derived in Lemma A.1 for this Φ satisfies $Z(\epsilon) = 0$ and $S''(M) = \Phi < 0$.

When $\theta < \eta$, Lemma A.4 implies that $Z_0(\epsilon)$ is positive and finite, and Lemma A.5 implies that $Z_\Phi(\epsilon)$ is negative and possibly equal to $-\infty$ for Φ positive and large. Hence, $\underline{\Phi} < 0$ and $\bar{\Phi} > 0$. If

$\bar{\Phi} < \infty$, then continuity and monotonicity of $Z_\Phi(\epsilon)$ in $[0, \bar{\Phi})$, $Z_0(\epsilon) > 0$ and $\lim_{\substack{\Phi \to \bar{\Phi} \\ \Phi < \bar{\Phi}}} Z_\Phi(\epsilon) = -\infty$ imply that there exists a unique $\Phi \in (0, \bar{\Phi})$ such that $Z_\Phi(\epsilon) = 0$. If $\bar{\Phi} = \infty$, then continuity and monotonicity of $Z_\Phi(\epsilon)$ in $[0, \infty)$, $Z_0(\epsilon) > 0$ and $\lim_{\Phi \to \infty} Z_\Phi(\epsilon) < 0$ imply that there exists a unique $\Phi \in (0, \infty)$ such that $Z_\Phi(\epsilon) = 0$. In both cases, there exists a unique $\Phi \in (0, \infty)$ such that $Z_\Phi(\epsilon) = 0$. The solution $S(D_t)$ derived in Lemma A.1 for this Φ satisfies $Z(\epsilon) = 0$ and $S''(M) = \Phi > 0$. □

Lemmas A.7-A.11 show properties of the solution derived in Lemma A.6. Lemma A.7 shows that the solution is increasing in D_t.

Lemma A.7. [Monotonicity and Positivity] *For the solution derived in Lemma A.6, $S(D_t) > 0$ and $S'(D_t) > 0$ for all $D_t \in [\epsilon, M]$.*

Proof of Lemma A.7. The solution derived in Lemma A.6 coincides with that derived in Lemma A.1 for a specific value of Φ. Hence, Lemma A.2 implies that either $S'(D_t) > 0$ for all D_t, or there exists $m < M$ such that $S'(D_t) > 0$ for all $D_t \in (m, M]$, $S'(D_t) < 0$ for all $D_t < m$, and $S(m) > 0$. In the second case, $S'(\epsilon) \leq 0$ and $S(\epsilon) > 0$, contradicting $Z(\epsilon) = 0$.

Since $S'(\epsilon) > 0$, $Z(\epsilon) = 0$ implies $S(\epsilon) > 0$. Combining $S(\epsilon) > 0$ with $S'(D_t) > 0$ for all D_t, we find $S(D_t) > 0$ for all D_t. □

Lemma A.8 compares the solution derived in Lemma A.6 to the affine solution derived in Proposition 3.1 for $L = \infty$. It shows that the former solution lies below the latter when $\theta > \eta$, and above it when $\theta < \eta$.

Lemma A.8. [Comparison with the affine solution for $L = \infty$] *Consider the solution derived in Lemma A.6, and the affine solution $a_0 + a_1 D_t$ derived in Proposition 3.1 for $L = \infty$. When $\theta > \eta$, $S(D_t) < a_0 + a_1 D_t$ for all $D_t \in [\epsilon, M]$, and when $\theta < \eta$, $S(D_t) > a_0 + a_1 D_t$ for all $D_t \in [\epsilon, M]$.*

Proof of Lemma A.8. We start with the case $\theta > \eta$, and consider the problem of maximizing

$$V(D_t) \equiv S(D_t) - (a_0 + a_1 D_t),$$

over the compact set $[\epsilon, M]$. The result in the lemma will follow if we show that the maximum value $V_{\max}$ of $V(D_t)$ is negative. Using (3.6), we can write $V(D_t)$ as

$$V(D_t) = S(D_t) - \frac{a_1}{r}(\kappa \bar{D} + r D_t).$$

Suppose first that $V(D_t)$ is maximized at $D_t = M$. Using (A.9), we can write $V(M)$ as

$$V(M) = \frac{1}{r}\left((\kappa\bar{D} + rM)(S'(M) - a_1) + \frac{1}{2}\sigma^2 M\Phi + \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{M}S'(M)\right). \tag{A.31}$$

Equations (3.7) and (A.8) imply that a_1 and $S'(M)$ are independent of M, and that $S'(M) < a_1$. Since, in addition $\Phi < 0$, (A.31) implies that $V_{\max} = V(M) < 0$ for M sufficiently large.

Suppose next that $V(D_t)$ is maximized at an interior point $m \in (\epsilon, M)$ that lies in the constrained region. The first- and second-order conditions of the maximization problem are $S'(m) = a_1$ and $S''(m) \le 0$. Setting $D_t = m$ in (3.13) and using $S'(m) = a_1$ and $S''(m) \le 0$, we find

$$\begin{aligned}
& m + \kappa(\bar{D} - m)a_1 - rS(m) \ge \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 m a_1^2 - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{m}a_1 \\
& \Rightarrow m + \kappa(\bar{D} - m)a_1 - rS(m) > \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 m a_1^2 - \frac{\rho(\theta-\eta)x}{1-x}\sigma^2 m a_1^2 \\
& \Leftrightarrow m + \kappa(\bar{D} - m)a_1 - rS(m) > \rho\theta\sigma^2 m a_1^2 \\
& \Leftrightarrow (\kappa\bar{D} + rm)a_1 - rS(m) > 0 \\
& \Leftrightarrow V_{\max} = V(m) < 0,
\end{aligned} \tag{A.32}$$

where the second step follows from (3.14) and the fourth step follows from (A.2).

Suppose next that $V(D_t)$ is maximized at an interior point $m \in (\epsilon, M)$ that lies in the unconstrained region. Setting $D_t = m$ in (3.4) and using $S'(m) = a_1$ and $S''(m) \le 0$, we find

$$\begin{aligned}
& m + \kappa(\bar{D} - m)a_1 - rS(m) \ge \rho\theta\sigma^2 m a_1^2 \\
& \Leftrightarrow (\kappa\bar{D} + rm)a_1 - rS(m) \ge 0 \\
& \Leftrightarrow V_{\max} = V(m) \le 0,
\end{aligned} \tag{A.33}$$

To show that (A.33) holds as a strict inequality, we proceed by contradiction. If (A.33) holds as an equality, then $S(m)$ and $S'(m)$ are the same as under the affine solution $S(D_t) = \frac{a_1}{r}(\kappa\bar{D} + rD_t)$. Hence, the solution derived in Lemma A.6 coincides with the affine solution in an interval in the unconstrained region that includes m and that has a boundary with the constrained region at an $m_1 \ge m$. Setting $D_t = m_1$ in (A.15) and using $S(m_1) = \frac{a_1}{r}(\kappa\bar{D} + rm_1)$, $S'(m_1) = a_1$ and $S''(m_1) = 0$, we find that the third derivative of $S(D_t)$ from the right at m_1 is

$$\frac{1}{2}\sigma^2 S'''(m_1) = \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2m_1^{\frac{3}{2}}}a_1 > 0. \tag{A.34}$$

Since $S'''(m_1) > 0$, $S''(D_t)$ is positive in a neigborhood to the right of m_1, and hence $S'(D_t)$ exceeds a_1. This means that $V(D_t)$, which is equal to zero for all $D_t \in [m, m_1]$ because $S(D_t)$ coincides with the affine solution, increases to the right of m_1, a contradiction since $V(D_t)$ would then be maximized in the constrained region.

If $V(D_t)$ is maximized at ϵ, then $S'(\epsilon) \leq a_1$ and hence,

$$V_{\max} = V(\epsilon) = S(\epsilon) - \frac{a_1}{r}(\kappa\bar{D} + r\epsilon) = \frac{1}{r}(\kappa\bar{D} + r\epsilon)(S'(\epsilon) - a_1) \leq 0, \tag{A.35}$$

where the second step follows from $Z(\epsilon) = 0$. To show that (A.35) holds as a strict inequality, we follow the same argument as in the case where $V(D_t)$ is maximized at an interior point m in the unconstrained region.

The argument in the case $\theta < \eta$ is symmetric. We consider the problem of minimizing $V(D_t)$ over $[\epsilon, M]$, and show that the minimum value $V_{\min}$ of $V(D_t)$ is positive.

Suppose first that $V(D_t)$ is minimized at $D_t = M$. Equations (3.7) and (A.8) imply that a_1 and $S'(M)$ are independent of M, and that $S'(M) > a_1$. Since, in addition, $\Phi > 0$, (A.31) implies that $V_{\min} = V(M) > 0$ for M sufficiently large.

Suppose next that $V(D_t)$ is maximized at an interior point $m \in (\epsilon, M)$ that lies in the constrained region. The first- and second-order conditions of the maximization problem are $S'(m) = a_1$ and $S''(m) \geq 0$. Setting $D_t = m$ in (3.13), using $S'(m) = a_1$ and $S''(m) \leq 0$, and proceeding as in the derivation of (A.32), we find $V_{\max} = V(m) > 0$.

Suppose next that $V(D_t)$ is maximized at an interior point $m \in (\epsilon, M)$ that lies in the unconstrained region. Setting $D_t = m$ in (3.4), using $S'(m) = a_1$ and $S''(m) \geq 0$, and proceeding as in the derivation of (A.33), we find $V_{\max} = V(m) \geq 0$. To show that (A.33) holds as a strict inequality, we follow the same argument as in the case $\theta > \eta$ and find that (A.34) implies $S'''(m_1) < 0$. This implies that $V(D_t)$, which is equal to zero for all $D_t \in [m, m_1]$, decreases to the right of m_1, a contradiction since $V(D_t)$ would then be minimized in the constrained region.

If $V(D_t)$ is maximized at ϵ, then $S'(\epsilon) \geq a_1$, and hence (A.35) implies $V_{\min} = V(\epsilon) \geq 0$. To show that (A.35) holds as a strict inequality, we follow the same argument as in the case where $V(D_t)$ is maximized at an interior point m in the unconstrained region.

Note that since $Z(\epsilon) = 0$ implies

$$S(\epsilon) - \frac{a_1}{r}(\kappa\bar{D} + r\epsilon) = \frac{1}{r}(\kappa\bar{D} + r\epsilon)(S'(\epsilon) - a_1),$$

Lemma A.8 implies that $S'(\epsilon) < a_1$ when $\theta > \eta$, and $S'(\epsilon) > a_1$ when $\theta < \eta$. □

Lemma A.9 shows that the constrained and the unconstrained regions have a single boundary and hence do not alternate. Proving this result requires condition $\kappa\bar{D} > \frac{\sigma^2}{4}$ of Theorem 3.1. This condition is required in all subsequent lemmas as well because they build on Lemma A.9, but is not used in all previous lemmas.

Lemma A.9. [Single boundary between unconstrained and constrained region] *There exists $m \in [\epsilon, M]$ such that the unconstrained region is $[\epsilon, m]$ and the constrained region is $(m, M]$.*

Proof of Lemma A.9. The constrained region includes a neighborhood to the left of M, for sufficiently large M, as shown in Lemma A.1. The unconstrained region includes a neighborhood to the right of ϵ, for sufficiently small ϵ. This is because $S'(\epsilon)$ is bounded above uniformly for all values of ϵ sufficiently small. When $\theta > \eta$, the upper bound is a_1. When $\theta < \eta$, Lemma A.5 implies that Φ is bounded above because otherwise $Z(\epsilon) < 0$. The upper bound on Φ implies one on $S(M)$ from (A.9), which in turn implies one on $S(\epsilon)$ from Lemma A.7, which in turn implies one on $S'(\epsilon)$ from $Z(\epsilon) = 0$.

Consider the non-empty set of $m > \epsilon$ such that $[\epsilon, m]$ lies in the unconstrained region, and the supremum m_1 of that set. Consider the non-empty set of $m > m_1$ such that (m_1, m) lies in the constrained region, and the supremum m_2 of that set. Suppose, by contradiction, that $m_2 < M$, in which case the unconstrained region begins again at m_2. Consider, in that case, the non-empty set of $m > \epsilon$ such that $[m_2, m]$ lies in the unconstrained region, and the supremum m_3 of that set. Since the constrained region includes a neighborhood to the left of M, $m_3 < M$.

Since (3.11) holds as an equality at m_i, $i = 1, 2, 3$,

$$\sqrt{m_i} S'(m_i) = \frac{L}{|\theta - \eta|\sigma} \equiv \hat{L}. \tag{A.36}$$

Since (3.11) holds to the left of m_i, $i = 1, 3$, and (3.14) holds to the right of m_i, the derivative of $\sqrt{D_t} S'(D_t)$ is non-negative for $D_t = m_i$, and hence

$$\sqrt{m} S''(m_i) + \frac{1}{2\sqrt{m_i}} S'(m_i) \geq 0 \Leftrightarrow m_i S''(m_i) \geq -\frac{S'(m_i)}{2} = -\frac{\hat{L}}{2} \frac{1}{\sqrt{m_i}} \quad \text{for} \quad i = 1, 3, \tag{A.37}$$

where the last step follows from (A.36). Conversely, since (3.14) holds to the left of m_2, and (3.11) holds to the right of m_2, the derivative of $\sqrt{D_t} S'(D_t)$ is non-positive for $D_t = m_2$, and hence

$$m_2 S''(m_2) \leq -\frac{S'(m_2)}{2} = -\frac{\hat{L}}{2} \frac{1}{\sqrt{m_2}}. \tag{A.38}$$

Since (3.14) holds in (m_1, m_2),

$$S(m_2) - S(m_1) = \int_{m_1}^{m_2} S'(D_t) dD_t > \int_{m_1}^{m_2} \frac{L}{|\theta - \eta|\sigma} \frac{1}{\sqrt{D_t}} dD_t = 2\hat{L}(\sqrt{m_2} - \sqrt{m_1}). \tag{A.39}$$

Conversely, since (3.11) holds in (m_2, m_3),

$$S(m_3) - S(m_2) = \int_{m_2}^{m_3} S'(D_t)dD_t \leq \int_{m_2}^{m_3} \frac{L}{|\theta-\eta|\sigma} \frac{1}{\sqrt{D_t}} dD_t = 2\hat{L}(\sqrt{m_3} - \sqrt{m_2}). \tag{A.40}$$

The points m_i, $i = 1, 2, 3$ satisfy (3.13) (as well as (3.4)). Setting $D_t = m_i$ in (3.13) and using (A.36), we find

$$m_i + \kappa(\bar{D} - m_i)\hat{L}\frac{1}{\sqrt{m_i}} + \frac{1}{2}\sigma^2 m_i S''(m_i) - rS(m_i) = \frac{\rho(\theta - x\eta)}{1-x}\sigma^2\hat{L}^2 - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\hat{L}. \tag{A.41}$$

Subtracting (A.41) for m_2 from the same equation for m_1, we find

$$\begin{aligned}
&m_1 - m_2 + \hat{L}\left[\kappa\bar{D}\left(\frac{1}{\sqrt{m_1}} - \frac{1}{\sqrt{m_2}}\right) - \kappa(\sqrt{m_1} - \sqrt{m_2})\right] \\
&+ \frac{1}{2}\sigma^2\left[m_1 S''(m_1) - m_2 S''(m_2)\right] - r[S(m_1) - S(m_2)] = 0, \\
&\Rightarrow m_1 - m_2 + \hat{L}\frac{m_2 - m_1}{\sqrt{m_1} + \sqrt{m_2}}\left(\kappa\bar{D}\frac{1}{\sqrt{m_1 m_2}} + \kappa\right) \\
&+ \hat{L}\frac{\sigma^2}{4}\left(\frac{1}{\sqrt{m_2}} - \frac{1}{\sqrt{m_1}}\right) + 2\hat{L}r(\sqrt{m_2} - \sqrt{m_1}) < 0 \\
&\Rightarrow \frac{\hat{L}}{\sqrt{m_1} + \sqrt{m_2}}\left(\frac{\kappa\bar{D} - \frac{\sigma^2}{4}}{\sqrt{m_1 m_2}} + \kappa + 2r\right) - 1 < 0,
\end{aligned} \tag{A.42}$$

where the second step follows from (A.37), (A.38) and (A.39), and the third step follows by dividing throughout by $m_2 - m_1 > 0$. Subtracting (A.41) for m_3 from the same equation for m_2, and using (A.37), (A.38) and (A.39), we similarly find

$$\frac{\hat{L}}{\sqrt{m_2} + \sqrt{m_3}}\left(\frac{\kappa\bar{D} - \frac{\sigma^2}{4}}{\sqrt{m_2 m_3}} + \kappa + 2r\right) - 1 \geq 0. \tag{A.43}$$

Condition $\kappa\bar{D} - \frac{\sigma^2}{4} > 0$ of Theorem 3.1 ensures that because $m_3 > m_1$, the left-hand side of (A.42) is larger than the left-hand side of (A.43). This is a contradiction because the former should be negative and the latter non-negative. Therefore, $m_2 = M$, and the lemma holds by setting $m = m_1$. □

Lemma A.10 shows that the solution is concave when $\theta > \eta$, and convex when $\theta < \eta$.

Lemma A.10. [Concavity/convexity] *The solution derived in Lemma A.6 satisfies $S''(D_t) < 0$ for all $D_t \in [\epsilon, M]$ when $\theta > \eta$, and $S''(D_t) > 0$ for all $D_t \in [\epsilon, M]$ when $\theta < \eta$.*

Proof of Lemma A.10. We start with the case $\theta > \eta$. Lemma A.6 shows that $S''(M) < 0$.

Moreover, setting $D_t = \epsilon$ in (3.4) and solving for $S''(\epsilon)$, we find

$$\begin{aligned}\frac{1}{2}\sigma^2\epsilon S''(\epsilon) &= \rho\theta\sigma^2\epsilon S'(\epsilon)^2 - \kappa(\bar{D}-\epsilon)S'(\epsilon) + rS(\epsilon) - \epsilon \\ &= \rho\theta\sigma^2\epsilon S'(\epsilon)^2 + (r+\kappa)\epsilon S'(\epsilon) - \epsilon \\ &= \epsilon\left(\rho\theta\sigma^2 S'(\epsilon)^2 + (r+\kappa)S'(\epsilon) - 1\right) < 0, \end{aligned} \tag{A.44}$$

where the second step follows from $Z(\epsilon) = 0$, and the last step because $S'(\epsilon) < a_1$.

Suppose, by contradiction, that there exists $D_t \in (\epsilon, M)$ such that $S''(D_t) \geq 0$, and consider the infimum m_1 within that set. Since $S''(\epsilon) < 0$, $m_1 > \epsilon$. The definition of m_1 implies $S''(D_t) < 0$ for all $D_t \in (\epsilon, m_1)$, $S''(m_1) = 0$ and $S'''(m_1) \geq 0$.

Suppose that m_1 lies in the unconstrained region. Setting $D_t = m_1$ in (A.18), and using $S''(m_1) = 0$, $S'''(m_1) \geq 0$ and (3.4), we find (A.19), written for m_1 instead of m. The contradiction follows because (i) $S''(D_t) < 0$ for all $D_t \in (\epsilon, m_1)$ implies $S'(m_1) < S'(\epsilon) < a_1$, (ii) the latter inequality together with $S'(m_1) > 0$ imply that the left-hand side of (A.2) becomes negative when a_1 is replaced by $S'(m_1)$.

Suppose next that m_1 lies in the constrained region and that $S'''(m_1) > 0$. Since $S''(m_1) = 0$, $S'''(m_1) > 0$ implies that $S''(D_t) > 0$ for D_t close to and larger than m_1. We denote by m_2 the supremum of the set of m such that $S''(D_t) > 0$ for all $D_t \in (m_1, m)$. Since $S''(M) < 0$, $m_2 < M$. The definition of m_2 implies $S''(D_t) > 0$ for all $D_t \in (m_1, m_2)$, $S''(m_2) = 0$ and $S'''(m_2) \leq 0$. Setting $D_t = m_1$ in (A.15), and using $S''(m_1) = 0$, $S'''(m_1) \geq 0$ and (3.13), we find (A.17), written for m_1 instead of m. Multiplying both sides by $\frac{m_1}{S'(m_1)} > 0$, we rewrite that equation as

$$-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2\sqrt{m_1}} + \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m_1) + r + \kappa - \frac{1}{S'(m_1)} \geq 0. \tag{A.45}$$

Since m_2 exceeds m_1, Lemma A.9 implies that it lies in the constrained region. Setting $D_t = m_2$ in (A.15), and using $S''(m_1) = 0$, $S'''(m_1) \leq 0$ and (3.13), we find (A.26), written for m_2 instead of m. Multiplying both sides by $\frac{m_2}{S'(m_2)} > 0$, we rewrite that equation as

$$-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{1}{2\sqrt{m_2}} + \frac{\rho(\theta - x\eta)}{1-x}\sigma^2 S'(m_2) + r + \kappa - \frac{1}{S'(m_2)} \leq 0. \tag{A.46}$$

Since $m_2 > m_1$ and $S'(m_2) > S'(m_1)$, the left-hand side of (A.46) is larger than the left-hand side of (A.45). This is a contradiction because the former should be non-positive and the latter non-negative.

Suppose finally that m_1 lies in the constrained region and that $S'''(m_1) = 0$. If there exists $D_t > m_1$ such that $S''(D_t) > 0$, then the same argument as in the case where $S'''(m_1) > 0$ yields a contradiction. If $S''(D_t) \leq 0$ for all $D_t > m_1$, then $S''(m_1) = S'''(m_1) = 0$ implies $S''''(m_1) \leq 0$. To derive a contradiction, we differentiate twice (3.13) at $D_t = m_1$. Using $S''(m_1) = S'''(m_1) = 0$, we find

$$\frac{1}{2}\sigma^2 m_1 S''''(m_1) = \frac{\rho \operatorname{sgn}(\theta - \eta) x L}{1 - x} \sigma \frac{1}{4 m_1^{\frac{3}{2}}} S'(m_1) > 0. \tag{A.47}$$

Hence, $S''(D_t) < 0$ for all $D_t \in [\epsilon, M]$.

We next consider the case $\theta < \eta$. Lemma A.6 shows that $S''(M) > 0$. Moreover, setting $D_t = \epsilon$ in (3.4), solving for $S''(\epsilon)$, and using $Z(\epsilon) = 0$, we find the following counterpart of (A.44)

$$\frac{1}{2}\sigma^2 \epsilon S''(\epsilon) = \epsilon \left(\rho \theta \sigma^2 S'(\epsilon)^2 + (r + \kappa) S'(\epsilon) - 1 \right). \tag{A.48}$$

When $\theta > 0$, (A.48) and $S'(\epsilon) > a_1$ imply $S''(\epsilon) > 0$. We next show the same result when $\theta < 0$. We rule out the cases $S'(\epsilon) < 0$ and $S'(\epsilon) = 0$ by contradiction arguments.

Suppose that $S''(\epsilon) < 0$. We denote by m_1 the supremum of the set of m such that $S''(D_t) < 0$ for all $D_t \in [\epsilon, m)$. Since $S''(M) > 0$, $m_1 < M$. The definition of m_1 implies $S''(D_t) < 0$ for all $D_t \in [\epsilon, m_1)$, $S''(m_1) = 0$ and $S'''(m_1) \geq 0$. Equations $Z(\epsilon) = 0$, (A.22) and $S''(D_t) < 0$ for all $D_t \in [\epsilon, m_1)$ imply $Z(m_1) < 0$.

If m_1 lies in the unconstrained region, (3.4) and $S''(m_1) = 0$ imply (A.24), written for m_1 instead of m. Moreover, setting $D_t = m_1$ in (A.18), and using $S''(m_1) = 0$, $S'''(m_1) \geq 0$ and (3.4), we find (A.19), written for m_1 instead of m. The two equations yield a contradiction when combined with $Z(m_1) < 0$.

If m_1 lies in the constrained region, (3.13) and $S''(m_1) = 0$ imply (A.27), written for m_1 instead of m. Moreover, setting $D_t = m_1$ in (A.15), and using $S''(m_1) = 0$, $S'''(m_1) \geq 0$ and (3.13), we find (A.17), written for m_1 instead of m. The two equations yield a contradiction when combined with $Z(m_1) < 0$, as can be seen by multiplying the latter equation by $-m_1^2$ and adding it to the former equation.

Suppose next that $S''(\epsilon) = 0$. Since $S'(\epsilon) > a_1$, (A.48) implies that $\theta < 0$ and $S'(\epsilon)$ is equal to the larger positive root of (A.2), which we denote by a_1^*. Hence, $S'(\epsilon)$ is the same as under the affine solution $S(D_t) = \frac{a_1^*}{r}(\kappa \bar{D} + r D_t)$. The same is true for $S(\epsilon)$ because of $Z(\epsilon) = 0$. Hence, the solution derived in Lemma A.6 coincides with the affine solution in an interval in the unconstrained region that includes ϵ and that has a boundary with the constrained region at an $m_1 \geq m$. Proceeding as in the proof of Lemma A.8, we find that the third derivative of $S(D_t)$ from the right at m_1 is negative, and hence $S''(D_t)$ is negative in a neigborhood to the right of m_1. Since $Z(m_1) = 0$, we can then use the previous argument to derive a contradiction. This establishes that $S''(\epsilon) > 0$.

Suppose, by contradiction, that there exists $D_t \in (\epsilon, M)$ such that $S''(D_t) \leq 0$, and consider the infimum m_1 within that set. Since $S''(\epsilon) > 0$, $m_1 > \epsilon$. The definition of m_1 implies $S''(D_t) > 0$ for all $D_t \in (\epsilon, m_1)$, $S''(m_1) = 0$ and $S'''(m_1) \leq 0$. Equations $Z(\epsilon) = 0$, (A.22) and $S''(D_t) > 0$ for all $D_t \in [\epsilon, m_1)$ imply $Z(m_1) > 0$.

Suppose that m_1 lies in the unconstrained region. Equations (3.4) and $S''(m_1) = 0$ imply (A.24), written for m_1 instead of m. Moreover, setting $D_t = m_1$ in (A.18), and using $S''(m_1) = 0$, $S'''(m_1) \leq 0$ and (3.4), we find (A.21), written for m_1 instead of m. The two equations yield a contradiction when combined with $Z(m_1) > 0$.

Suppose next that m_1 lies in the constrained region and that $S'''(m_1) < 0$. Since $S''(m_1) = 0$, $S'''(m_1) < 0$ implies that $S''(D_t) < 0$ for D_t close to and larger than m_1. We denote by m_2 the supremum of the set of m such that $S''(D_t) < 0$ for all $D_t \in (m_1, m)$. Since $S''(M) > 0$, $m_2 < M$. The definition of m_2 implies $S''(D_t) < 0$ for all $D_t \in (m_1, m_2)$, $S''(m_2) = 0$ and $S'''(m_2) \geq 0$. Setting $D_t = m_1$ in (A.15), and using $S''(m_1) = 0$, $S'''(m_1) \leq 0$ and (3.13), we find (A.26), written for m_1 instead of m. Multiplying both sides by m_1^2, we rewrite that equation as

$$-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{\sqrt{m_1}}{2}S'(m_1) + m_1\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m_1)^2 + (r+\kappa)S'(m_1) - 1\right) \leq 0. \quad \text{(A.49)}$$

Since m_2 exceeds m_1, Lemma A.9 implies that it lies in the constrained region. Setting $D_t = m_2$ in (A.15), and using $S''(m_1) = 0$, $S'''(m_1) \geq 0$ and (3.13), we find (A.17), written for m_2 instead of m. Multiplying both sides by m_2^2, we rewrite that equation as

$$-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{\sqrt{m_2}}{2}S'(m_2) + m_2\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m_2)^2 + (r+\kappa)^2 S'(m_2) - 1\right) \geq 0. \quad \text{(A.50)}$$

Since $S''(D_t) < 0$ for all $D_t \in (m_1, m_2)$, $Z(m_2) < Z(m_1)$ Using (A.27) to compute $Z(m_1)$ and $Z(m_2)$, we find

$$\begin{aligned} &m_1\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m_1)^2 + (r+\kappa)S'(m_1) - 1\right) - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{m_1}S'(m_1) \\ &> m_2\left(\frac{\rho(\theta-x\eta)}{1-x}\sigma^2 S'(m_2)^2 + (r+\kappa)S'(m_2) - 1\right) - \frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\sqrt{m_2}S'(m_2). \end{aligned} \quad \text{(A.51)}$$

Multiplying (A.49) by -1 and adding to the sum of (A.50) and (A.51), we find

$$-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{\sqrt{m_1}}{2}S'(m_1) > -\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\sigma\frac{\sqrt{m_2}}{2}S'(m_2),$$

and hence $\sqrt{m_1}S'(m_1) > \sqrt{m_2}S'(m_2)$. Consider the value $\hat{m}$ that minimizes the function $\sqrt{D_t}S'(D_t)$ over the compact set $[m_1, M]$. Since $\sqrt{m_1}S'(m_1) > \sqrt{m_2}S'(m_2)$, $\hat{m} > m_1$. Since $S''(M) > 0$, the function $\sqrt{D_t}S'(D_t)$ is increasing for D_t close to and smaller than M, and hence $\hat{m} < M$. Since

$\hat{m}$ is an interior minimum and not all n-order derivatives of $\sqrt{D_t}S'(D_t)$ are zero at $\hat{m}$ (proof of Lemma A.1), the smallest n for which the n-order derivative is non-zero is even. Since $\hat{m}$ is in the constrained region, $\sqrt{\hat{m}}S'(\hat{m}) \geq \frac{L}{|\theta-\eta|\sigma}$. We can hence choose $\hat{m}_1 < \hat{m}_2 < \hat{m}_3$ in the constrained region such that (i) $\hat{m}_2 < \hat{m} < \hat{m}_3$, (ii) $\sqrt{\hat{m}_1}S'(\hat{m}_1) = \sqrt{\hat{m}_2}S'(\hat{m}_2) = \sqrt{\hat{m}_3}S'(\hat{m}_3) \equiv \hat{L} > \frac{L}{|\theta-\eta|\sigma}$, (iii) $\sqrt{D_t}S'(D_t) > \hat{L}$ for all $D_t \in (\hat{m}_1, \hat{m}_2)$, and (iv) $\sqrt{D_t}S'(D_t) < \hat{L}$ for all $D_t \in (\hat{m}_2, \hat{m}_3)$. We can then proceed as in the proof of Lemma A.9 to derive a contradiction. (In the proof of Lemma A.9, $\hat{L} = \frac{L}{|\theta-\eta|\sigma}$, but the proof works for any value of $\hat{L}$.)

Suppose finally that m_1 lies in the constrained region and that $S'''(m_1) = 0$. If there exists $D_t > m_1$ such that $S''(D_t) < 0$, then the same argument as in the case where $S'''(m_1) < 0$ yields a contradiction. If $S''(D_t) \geq 0$ for all $D_t > m_1$, then $S''(m_1) = S'''(m_1) = 0$ implies $S''''(m_1) \geq 0$. To derive a contradiction, we differentiate twice (3.13) at $D_t = m_1$. Using $S''(m_1) = S'''(m_1) = 0$, we find

$$\frac{1}{2}\sigma^2 m_1 S''''(m_1) = \frac{\rho \operatorname{sgn}(\theta-\eta) x L}{1-x}\sigma \frac{1}{4m_1^{\frac{3}{2}}} S'(m_1) < 0. \tag{A.52}$$

Hence, $S''(D_t) > 0$ for all $D_t \in [\epsilon, M]$. □

Lemma A.11 completes the comparison of the solution derived in Lemma A.6 to the affine solution derived in Proposition 3.1 for $L = \infty$. It shows that the derivative of the former solution lies below the derivative of the latter when $\theta > \eta$, and above it when $\theta < \eta$.

Lemma A.11. [Comparison with the derivative of the affine solution] *Consider the solution derived in Lemma A.6, and the affine solution $a_0 + a_1 D_t$ derived in Proposition 3.1 for $L = \infty$. When $\theta > \eta$, $S'(D_t) < a_1$ for all $D_t \in [\epsilon, M]$, and when $\theta < \eta$, $S'(D_t) > a_1$ for all $D_t \in [\epsilon, M]$.*

Proof of Lemma A.11. When $\theta > \eta$, the result follows because the solution is concave and $S'(\epsilon) < a_1$. When $\theta < \eta$, the result follows because the solution is convex and $S'(\epsilon) > a_1$. □

Lemma A.12 shows that the comparisons in Lemmas A.8 and A.11 are reversed when they concern the affine solution derived in Proposition 3.2 for $L = 0$.

Lemma A.12. [Comparison with the affine solution for $L = 0$] *Consider the solution derived in Lemma A.6, and the affine solution $a_0 + a_1 D_t$ derived in Proposition 3.2 for $L = 0$. When $\theta > \eta$, $S(D_t) > a_0 + a_1 D_t$ and $S'(D_t) > a_1$ for all $D_t \in [\epsilon, M)$. When $\theta < \eta$, $S(D_t) < a_0 + a_1 D_t$ and $S'(D_t) < a_1$ for all $D_t \in [\epsilon, M)$.*

Proof of Lemma A.12. When $\theta > \eta$, $S'(M) = a_1$ and concavity of $S(D_t)$ imply $S'(D_t) > a_1$ for all $D_t \in [\epsilon, M)$. Hence, the function

$$V(D_t) \equiv S(D_t) - (a_0 + a_1 D_t)$$

is increasing. That function is also positive because

$$V(\epsilon) = S(\epsilon) - (a_0 + a_1\epsilon) = \frac{1}{r}(\kappa\bar{D} + r\epsilon)(S'(\epsilon) - a_1) > 0,$$

where the second step follows from $Z(\epsilon) = 0$ and (3.6). Therefore, $S(D_t) > a_0 + a_1 D_t$ for all $D_t \in [\epsilon, M]$.

When $\theta < \eta$, $S'(M) = a_1$ and convexity of $S(D_t)$ imply $S'(D_t) < a_1$ for all $D_t \in [\epsilon, M)$. Hence, the function $V(D_t)$ is decreasing. That function is also negative because

$$V(\epsilon) = \frac{1}{r}(\kappa\bar{D} + r\epsilon)(S'(\epsilon) - a_1) < 0.$$

Therefore, $S(D_t) < a_0 + a_1 D_t$ for all $D_t \in [\epsilon, M]$. □

Lemma A.13 shows that if a solution to the system of ODEs exists in $(0, \infty)$ and its derivative converges to finite limits at zero and infinity, then these limits are almost uniquely determined.

Lemma A.13. [Limits at zero and infinity] *Consider a solution $S(D_t)$ to the system of ODEs (3.4) in the unconstrained region (3.11), and (3.13) in the constrained region (3.14), defined in $(0, \infty)$. Suppose that $S'(D_t)$ converges to finite limits at zero and infinity, denoted by $S'(0)$ and $S'(\infty)$, respectively. Then $S'(\infty)$ is a root of (A.11), and $S'(0)$ satisfies $Z(0) \equiv \kappa\bar{D}S'(0) - rS(0) = 0$, where $S(0)$ denotes the limit of $S(D_t)$ at zero.*

Proof of Lemma A.13. We start with the limit at zero. Since $\lim_{D_t\to 0} S'(D_t)$ exists and is finite, the same is true for $\lim_{D_t\to 0} S(D_t)$. (The latter limit is $S(D_t) - \int_0^{D_t} S'\left(\hat{D}_t\right) d\hat{D}_t$ for any given D_t).

Since $\lim_{D_t\to 0} S'(D_t)$ exists and is finite, values of D_t close to zero lie in the unconstrained region. Moreover, since $\lim_{D_t\to 0} S'(D_t)$ and $\lim_{D_t\to 0} S(D_t)$ exist and are finite, taking the limit of both sides of the ODE (3.4) when D_t goes to zero implies that $\lim_{D_t\to 0} D_t S''(D_t)$ exists and is finite. If the latter limit differs from zero, then $|S''(D_t)| \geq \frac{\ell}{D_t}$ for $\ell > 0$ and for all D_t smaller than a sufficiently small $\eta > 0$. Since, however, for $D_t < \eta$,

$$S'(D_t) = S'(\eta) + \int_\eta^{D_t} S''(\hat{D}_t) d\hat{D}_t \Rightarrow |S'(D_t) - S'(\eta)| \geq \int_{D_t}^\eta \frac{\ell}{\hat{D}_t} d\hat{D}_t = \ell \log\left(\frac{\eta}{D_t}\right),$$

$\lim_{D_t\to 0} S'(D_t)$ is plus or minus infinity, a contradiction. Hence, $\lim_{D_t\to 0} D_t S''(D_t) = 0$. Taking the limit of (3.4) when D_t goes to zero, and using $\lim_{D_t\to 0} S'(D_t) = S'(0)$, $\lim_{D_t\to 0} S'(D_t) = S(0)$, $\lim_{D_t\to 0} D_t S''(D_t) = 0$, and the finiteness of $S'(0)$ and $S(0)$, we find $Z(0) = 0$.

We next consider the limit at infinity. Since $\lim_{D_t\to\infty} S'(D_t)$ exists and is finite, it is equal to $\lim_{D_t\to\infty} \frac{S(D_t)}{D_t}$. This follows by writing $\frac{S(D_t)}{D_t}$ as

$$\frac{S(D_t)}{D_t} = \frac{S(0) + \int_0^{D_t} S'\left(\hat{D}_t\right) d\hat{D}_t}{D_t},$$

and noting that $\lim_{D_t\to\infty} \frac{S(0)}{D_t} = 0$ and $\lim_{D_t\to\infty} \frac{\int_0^{D_t} S'(\hat{D}_t) d\hat{D}_t}{D_t} = \lim_{D_t\to\infty} S'(D_t)$.

Since $\lim_{D_t\to\infty} S'(D_t)$ exists and is finite, large values of D_t lie in the constrained region. Dividing both sides of the ODE (3.13) by D_t, taking the limit when D_t goes to infinity, and using the existence and finiteness of $\lim_{D_t\to\infty} S'(D_t)$ and $\lim_{D_t\to\infty} \frac{S(D_t)}{D_t}$, we find that $\lim_{D_t\to\infty} S''(D_t)$ exists and is finite. If the latter limit differs from zero, then $|S''(D_t)| \geq \ell > 0$ for $\ell > 0$ and for all D_t sufficiently large, implying that $\lim_{D_t\to 0} S'(D_t)$ is plus or minus infinity, a contradiction. Taking the limit of (3.13) divided by D_t when D_t goes to infinity, and using $\lim_{D_t\to\infty} S'(D_t) = \lim_{D_t\to 0} \frac{S(D_t)}{D_t} = S'(\infty)$, $\lim_{D_t\to 0} S''(D_t) = 0$, and the finiteness of $S'(\infty)$, we find that $S'(\infty)$ is a root of (A.11). □

Lemma A.14 shows that a solution to the system of ODEs with a derivative that converges to finite limits at zero and infinity exists in $(0,\infty)$, and has the properties in Lemmas A.7-A.11.

Lemma A.14. [Existence in $(0,\infty)$] *A solution $S(D_t)$ to the system of ODEs (3.4) in the unconstrained region (3.11), and (3.13) in the constrained region (3.14), with a derivative that converges to finite limits at zero and infinity exists in $(0,\infty)$, and has the properties in Lemmas A.7-A.12.*

Proof of Lemma A.14. We will construct the solution in $(0,\infty)$ as the simple limit of solutions in compact intervals $[\epsilon, M]$. We denote by $S_{\epsilon,M}(D_t)$ the solution derived in Lemma A.6, and by $\Phi_{\epsilon,M}$ and $Z_{\epsilon,M}(D_t)$ the corresponding values of Φ and $Z(D_t)$.

We start with the case $\theta > \eta$, and first derive the limit when ϵ goes to zero, holding M constant. Consider $\epsilon_1 > \epsilon_2 > 0$, and suppose, by contradiction, that $\Phi_{\epsilon_2,M} \leq \Phi_{\epsilon_1,M}$. Lemma A.3 then implies $S_{\epsilon_2,M}(\epsilon_1) \leq S_{\epsilon_1,M}(\epsilon_1)$ and $S'_{\epsilon_2,M}(\epsilon_1) \geq S'_{\epsilon_1,M}(\epsilon_1)$, which in turn imply $Z_{\epsilon_2,M}(\epsilon_1) \geq Z_{\epsilon_1,M}(\epsilon_1) = 0$. This is a contradiction because $S''_{\epsilon_2,M}(D_t) < 0$ and $Z_{\epsilon_2,M}(\epsilon_2) = 0$ imply $Z_{\epsilon_2,M}(\epsilon_1) < 0$. Hence, $\Phi_{\epsilon_2,M} > \Phi_{\epsilon_1,M}$, and Lemma A.3 implies $S_{\epsilon_2,M}(D_t) > S_{\epsilon_1,M}(D_t)$ and $S'_{\epsilon_2,M}(D_t) < S'_{\epsilon_1,M}(D_t)$ for all $D_t \in (\epsilon_1, M)$.

Since for given $D_t \in (0, M)$, the function $\epsilon \to S_{\epsilon,M}(D_t)$, defined for $\epsilon < D_t$, increases as

ϵ decreases and is bounded above by the affine solution derived for $L = \infty$ (Lemma A.8), it converges to a finite limit $S_M(D_t)$ when ϵ goes to zero. Likewise, since for given D_t, the function $\epsilon \to S'_{\epsilon,M}(D_t)$, defined for $\epsilon < D_t$, decreases as ϵ decreases and is bounded below by zero (Lemma A.7), it converges to a finite limit $\hat{S}_M(D_t)$ when ϵ goes to zero.

The simple limit $S_M(D_t)$ of $S_{\epsilon,M}(D_t)$ is differentiable, and its derivative is the simple limit $\hat{S}_M(D_t)$ of $S'_{\epsilon,M}(D_t)$. To show this result, we use the intermediate value theorem together with a uniform bound on $S''_{\epsilon,M}(D_t)$. The function $S_{\epsilon,M}(D_t)$ is bounded above by the affine solution $\frac{a_1}{r}(\kappa\bar{D} + rD_t)$ and below by zero (Lemma A.7). Likewise, the function $S'_{\epsilon,M}(D_t)$ is bounded above by a_1 (Lemma A.11) and below by zero. Hence, for any given D_t and neighborhood $\mathcal{N}$ around D_t, the ODEs (3.4) and (3.13) imply a bound Q on $S''_{\epsilon,M}(m)$ that is uniform over $m \in \mathcal{N}$, ϵ and M. The intermediate value theorem implies that for $m \in \mathcal{N}$,

$$\left|\frac{S_{\epsilon,M}(m) - S_{\epsilon,M}(D_t)}{m - D_t} - S'_{\epsilon,M}(D_t)\right| = \left|S'_{\epsilon,M}(m') - S'_{\epsilon,M}(D_t)\right| = \left|S''_{\epsilon,M}(m'')\right| |m' - D_t| < Q|m - D_t|,$$

where m' is between m and D_t, and m'' is between m' and D_t. Taking the limit when ϵ goes to zero, we find

$$\left|\frac{S_M(m) - S_M(D_t)}{m - D_t} - \hat{S}_M(D_t)\right| \leq Q|m - D_t|,$$

which establishes that $S_M(D_t)$ is differentiable at D_t and its derivative is $S'_M(D_t) = \hat{S}_M(D_t)$. Since $S'_{\epsilon,M}(D_t)$ and $S'_{\epsilon,M}(D_t)$ have simple limits, we can use the ODEs (3.4) and (3.13) to construct a simple limit for $S''_{\epsilon,M}(D_t)$, which we denote by $\hat{\hat{S}}_M(D_t)$. The same argument that establishes $S'_M(D_t) = \hat{S}_M(D_t)$ can be used to establish $\hat{\hat{S}}_M(D_t) = S''_M(D_t)$, and hence that $S_M(D_t)$ solves the system of ODEs in $(0, M]$. Since $S'_{\epsilon,M}(D_t)$ is decreasing in D_t and is bounded below by zero, its limit $S'_M(D_t)$ over ϵ is non-increasing in D_t and has the same lower bound. Hence, $S'_M(D_t)$ converges to a finite limit $S'_M(0)$ when D_t goes to zero. Using the same argument as in Lemma A.13, we can show that $Z_M(0) \equiv \kappa\bar{D}S'_M(0) - rS_M(0) = 0$, where $S_M(0)$ denotes the limit of $S_M(D_t)$ when D_t goes to zero.

Since $S_M(D_t)$, $S'_M(D_t)$ and $S''_M(D_t)$ are the simple limits of $S_{\epsilon,M}(D_t)$, $S'_{\epsilon,M}(D_t)$ and $S''_{\epsilon,M}(D_t)$, respectively, the properties in Lemmas A.7, A.8, A.10, A.11 and A.12 hold as weak inequalities for all $D_t \in (0, M]$. Following similar arguments as in these Lemmas and using $Z_M(0) = 0$, we can show that the inequalities are strict.

We next derive the limit when M goes to infinity. Consider $M_2 > M_1$. Since $S''_{M_2}(D_t) < 0$ and $S'_{M_2}(M_1) = S'_{M_1}(M_1)$, $S'_{M_2}(M_1) > S'_{M_1}(M_1)$. Suppose, by contradiction, that $S_{M_2}(M_1) \leq S_{M_1}(M_1)$. Equations $S_{M_2}(M_1) \leq S_{M_1}(M_1)$ and $S'_{M_2}(M_1) > S'_{M_1}(M_1)$ imply $S_{M_2}(D_t) < S_{M_1}(D_t)$

for D_t smaller than and close to M_1. The same argument as in Lemma A.3 then implies $S_{M_2}(D_t) < S_{M_1}(D_t)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ for all $D_t \in (0, M_1)$. Since $S_{M_2}(M_1) \leq S_{M_1}(M_1)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ for all $D_t \in (0, M_1)$, $S_{M_2}(0) < S_{M_1}(0)$. Combining the latter equation with $S'_{M_2}(0) \geq S'_{M_1}(0)$, which follows by taking the limit of $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ when D_t goes to zero, we find $Z_{M_2}(0) > Z_{M_1}(0)$, a contradiction since $Z_{M_2}(0) = Z_{M_1}(0) = 0$. Hence, $S_{M_2}(M_1) > S_{M_1}(M_1)$.

The inequalities $S_{M_2}(D_t) > S_{M_1}(D_t)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ hold by continuity for D_t smaller than and close to M_1. Suppose, by contradiction, that there exists $D_t \in (0, M_1)$ such that $S_{M_2}(D_t) \leq S_{M_1}(D_t)$ or $S'_{M_2}(D_t) \leq S'_{M_1}(D_t)$, and consider the supremum m within that set. The definition of m implies $S_{M_2}(D_t) > S_{M_1}(D_t)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ for all $D_t \in (m, M_1)$, and $S_{M_2}(m) = S_{M_1}(m)$ or $S'_{M_2}(m) = S'_{M_1}(m)$. Only one of the latter two equations holds since otherwise the solutions $S_{M_1}(D_t)$ and $S_{M_2}(D_t)$ would coincide. If $S_{M_2}(m) = S_{M_1}(m)$ and $S'_{M_2}(m) > S'_{M_1}(m)$, then $S_{M_2}(D_t) < S_{M_1}(D_t)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ for D_t smaller than and close to M_1. The same argument as in Lemma A.3 then implies $S_{M_2}(D_t) < S_{M_1}(D_t)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ for all $D_t \in (0, m)$. This, in turn, implies $Z_{M_2}(0) > Z_{M_1}(0)$, a contradiction. If instead $S_{M_2}(m) > S_{M_1}(m)$ and $S'_{M_2}(m) = S'_{M_1}(m)$, then the same argument as in Lemma A.3 implies $S_{M_2}(D_t) > S_{M_1}(D_t)$ and $S'_{M_2}(D_t) < S'_{M_1}(D_t)$ for all $D_t \in (0, m)$. This, in turn, implies $Z_{M_2}(0) < Z_{M_1}(0)$, a contradiction. Hence, $S_{M_2}(D_t) > S_{M_1}(D_t)$ and $S'_{M_2}(D_t) > S'_{M_1}(D_t)$ for all $D_t \in (0, M_1)$.

Since for given $D_t \in (0, \infty)$, the function $M \to S_M(D_t)$, defined for $D_t < M$, is increasing in M and is bounded above by the affine solution derived for $L = \infty$, it converges to a finite limit $S(D_t)$ when M goes to infinity. Likewise, since for given $D_t \in (0, \infty)$, the function $M \to S'_M(D_t)$, defined for $D_t < M$, is increasing in M and is bounded above by a_1, it converges to a finite limit $\hat{S}(D_t)$ when M goes to infinity. The same argument as when taking the limit over ϵ establishes that $\hat{S}(D_t) = S'(D_t)$ and that $S(D_t)$ solves the system of ODEs in $(0, \infty)$. Since $S'_M(D_t)$ is decreasing in D_t and is bounded below by zero and above by a_1, its limit $S'(D_t)$ over M is non-increasing in D_t and has the same bounds. Hence, $S'(D_t)$ converges to finite limits $S'(0)$ when D_t goes to zero and $S'(\infty)$ when D_t goes to infinity. Lemma A.13 implies that $Z(0) \equiv \kappa \bar{D} S'(0) - rS(0) = 0$, where $S(0)$ denotes the limit of $S(D_t)$ when D_t goes to zero. Lemma A.13 also implies that $S'(\infty)$ is a root of (A.11). Since $S'(D_t)$ is the simple limit of $S'_M(D_t)$, which is positive and increasing in M, it is positive. Hence, $S'(\infty)$ is non-negative and equal to the unique positive root of (A.11). The same arguments as when taking the limit over ϵ establish that the properties in Lemmas A.7, A.8, A.10, A.11 and A.12 hold for all $D_t \in (0, \infty)$.

The argument in the case $\theta < \eta$ is symmetric. The monotonicity of $S_{\epsilon,M}(D_t)$ and $S'_{\epsilon,M}(D_t)$ as functions of ϵ, and of $S_M(D_t)$ and $S'_M(D_t)$ as functions of M, is the opposite relative to the case $\theta > \eta$. When $\theta < x\eta$, the limit $S'(\infty)$ is equal to the smaller of the two positive roots of (A.11) because $S'(D_t)$ is bounded above by that root. The upper bound on $S'(D_t)$ follows from the same upper bound on $S'_M(D_t)$: convexity implies that $S'_M(D_t) < S'_M(M)$ for all $D_t \in (0, M)$, and $S'_M(M)$ is equal to the smaller positive root of (A.11). □

Theorem 3.1 follows from Lemma A.14. □

Proof of Proposition 3.3. Substituting the asset's share return from (3.1) into (2.2), and setting $S_t = S(D_t)$, we find that the asset's dollar return is

$$dR_t = \frac{\left[D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t)\right] dt + \sigma\sqrt{D_t}S'(D_t)dB_t}{S(D_t)} - rdt. \tag{A.53}$$

The return's conditional volatility is the diffusion coefficient in (A.53) times $\sqrt{dt}$:

$$\sqrt{\mathbb{V}\mathrm{ar}_t(dR_t)} = \frac{\sigma\sqrt{D_t}S'(D_t)\sqrt{dt}}{S(D_t)}. \tag{A.54}$$

The return's conditional volatility under the affine solutions derived for $L = 0$ and $L = \infty$ is given by (A.5). Comparing (A.5) and (A.54), we find that the return's conditional volatility is higher than under the affine solutions if

$$S'(D_t)(\kappa\bar{D} + rD_t) > rS(D_t) \Leftrightarrow Z(D_t) > 0,$$

and is lower than under the affine solutions if $Z(D_t) < 0$. When $\theta > \eta$, $Z(0) = 0$ and concavity imply $Z(D_t) < 0$, and hence conditional volatility is lower than under the affine solutions. When instead $\theta < \eta$, $Z(0) = 0$ and convexity imply $Z(D_t) > 0$, and hence conditional volatility is higher than under the affine solutions. The comparison of conditional volatility across the cases $\theta > \eta$ and $\theta < \eta$ follows from the comparison of each case with the affine solutions since volatility under the affine solutions is independent of θ.

Since the return's unconditional variance is the unconditional expectation of the return's conditional variance, the comparisons derived for conditional volatility carry over to unconditional volatility. □

Proof of Proposition 3.4. The conditional beta of asset n is

$$\beta_{nt} = \frac{\mathbb{C}\mathrm{ov}_t(dR_{nt}, dR_{Mt})}{\mathbb{V}\mathrm{ar}_t(dR_{Mt})}, \tag{A.55}$$

where dR_{nt} denotes the return of asset n and dR_{Mt} denotes the return of the market portfolio.

Assuming that the market portfolio includes ζ_m shares of asset $m = 1, .., N$, its return is

$$dR_{Mt} = \frac{dR^{sh}_{Mt}}{S_{Mt}} = \frac{\sum_{m=1}^{N} \zeta_m dR^{sh}_{mt}}{\sum_{m=1}^{N} \zeta_m S_{mt}} = \sum_{m=1}^{N} \frac{\zeta_m S_{mt}}{\sum_{m=1}^{N} \zeta_m S_{mt}} dR_{mt} = \sum_{m=1}^{N} \omega_{mt} dR_{mt}, \tag{A.56}$$

where S_{Mt} denotes the market portfolio's price and

$$\omega_{mt} \equiv \frac{\zeta_m S_{mt}}{\sum_{m=1}^{N} \zeta_m S_{mt}}$$

denotes the weight of asset n in the market portfolio. Equation (A.55) implies that the conditional beta of asset n exceeds that of asset n' if

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}_t(dR_{nt}, dR_{Mt}) > \mathbb{C}\mathrm{ov}_t(dR_{n't}, dR_{Mt}) \\
&\Leftrightarrow \omega_n \mathbb{V}\mathrm{ar}_t(dR_{nt}) > \omega_{n'} \mathbb{V}\mathrm{ar}_t(dR_{n't}) \\
&\Leftrightarrow \zeta_n S_{nt} \mathbb{V}\mathrm{ar}_t(dR_{nt}) > \zeta_{n'} S_{n't} \mathbb{V}\mathrm{ar}_t(dR_{n't}),
\end{aligned} \tag{A.57}$$

where the second step follows from (A.56) and the independence of returns across assets.

Suppose next that $\theta_n < \eta < \theta_{n'}$, and that other characteristics of assets n and n' are identical ($\kappa_n = \kappa_{n'}$, $\bar{D}_n = \bar{D}_{n'}$, $\sigma_n = \sigma_{n'}$, $D_{nt} = D_{n't}$ and $\zeta_n = \zeta_{n'}$). Since a_1 decreases in θ (Proposition 3.1), the affine solution derived for $L = \infty$ is larger for θ_n than for θ'_n. Since, in addition, S_{nt} lies above the affine solution for θ_n, while $S_{n't}$ lies below the affine solution for θ'_n (Theorem 3.1), $S_{nt} > S_{n't}$. Since, finally, $\mathbb{V}\mathrm{ar}_t(dR_{nt}) > \mathbb{V}\mathrm{ar}_t(dR_{n't})$ (Proposition 3.3), (A.57) implies $\mathbb{C}\mathrm{ov}_t(dR_{nt}, dR_{Mt}) > \mathbb{C}\mathrm{ov}_t(dR_{n't}, dR_{Mt})$ and hence $\beta_{nt} > \beta_{n't}$.

The unconditional beta of asset n is

$$\beta_{nt} = \frac{\mathbb{C}\mathrm{ov}(dR_{nt}, dR_{Mt})}{\mathbb{V}\mathrm{ar}(dR_{Mt})} = \frac{\mathbb{E}\left(\mathbb{C}\mathrm{ov}_t(dR_{nt}, dR_{Mt})\right)}{\mathbb{E}\left(\mathbb{V}\mathrm{ar}_t(dR_{Mt})\right)},$$

Since the conditional covariance of $\mathbb{C}\mathrm{ov}_t(dR_{nt}, dR_{Mt})$ is larger for asset n than for asset n', the same is true for the unconditional covariance, and hence for the unconditional beta. □

Proof of Proposition 3.5. Since (3.6) implies $S(D_t) = a_1 \left(\frac{\kappa}{r}\bar{D} + D_t\right)$, (3.17) is equivalent to

$$a_{1n} + a_{1n'} - (a_{1\hat{n}} + a_{1\hat{n}'}) > 0. \tag{A.58}$$

When $L = \infty$, Proposition 3.1 implies that (A.58) is equivalent to

$$\Psi(\theta_n) + \Psi(\theta_{n'}) - [\Psi(\theta_{\hat{n}}) + \Psi(\theta_{\hat{n}'})] > 0, \tag{A.59}$$

where the function $\Psi(\theta)$ is defined in the proof of Proposition 3.1. Setting $\ell \equiv \bar{\theta} - \theta_n = \theta_{n'} - \bar{\theta} > 0$ and $\hat{\ell} \equiv \bar{\theta} - \theta_{\hat{n}} = \theta_{\hat{n}'} - \bar{\theta} \in (0, \ell)$, we can write (A.59) as

$$\begin{aligned}
&\Psi(\bar{\theta} - \ell) + \Psi(\bar{\theta} + \ell) - \left[\Psi(\bar{\theta} - \hat{\ell}) + \Psi(\bar{\theta} + \hat{\ell})\right] > 0 \\
&\int_{\hat{\ell}}^{\ell} \Psi'(\bar{\theta} + x)dx - \int_{\hat{\ell}}^{\ell} \Psi'(\bar{\theta} - x)dx > 0 \\
&\Leftrightarrow \int_{\hat{\ell}}^{\ell} \left(\int_{-x}^{x} \Psi''(\bar{\theta} + y)dy\right) dx > 0. \qquad \text{(A.60)}
\end{aligned}$$

Equation (A.60) holds because $\Psi(\theta)$ is convex. When $L = 0$, Proposition 3.2 implies that (A.58) is equivalent to (A.59) with the function $\Psi(\frac{\theta - x\eta}{1-x})$ instead of $\Psi(\theta)$. Since $\Psi(\frac{\theta - x\eta}{1-x})$ is convex, the modified (A.59) holds.

Propositions 3.1 and 3.2 imply that the comparison between $L = \infty$ and $L = 0$ in the corollary is equivalent to

$$\begin{aligned}
&\Psi\left(\frac{\theta_n - x\eta}{1-x}\right) + \Psi\left(\frac{\theta_{n'} - x\eta}{1-x}\right) - \left[\Psi\left(\frac{\theta_{\hat{n}} - x\eta}{1-x}\right) + \Psi\left(\frac{\theta_{\hat{n}'} - x\eta}{1-x}\right)\right] \\
&> \Psi(\theta_n) + \Psi(\theta_{n'}) - [\Psi(\theta_{\hat{n}}) + \Psi(\theta_{\hat{n}'})] \\
&\Leftrightarrow \Psi\left(\frac{\bar{\theta} - x\eta}{1-x} - \frac{\ell}{1-x}\right) + \Psi\left(\frac{\bar{\theta} - x\eta}{1-x} + \frac{\ell}{1-x}\right) - \left[\Psi\left(\frac{\bar{\theta} - x\eta}{1-x} - \frac{\hat{\ell}}{1-x}\right) + \Psi\left(\frac{\bar{\theta} - x\eta}{1-x} + \frac{\hat{\ell}}{1-x}\right)\right] \\
&> \Psi(\bar{\theta} - \ell) + \Psi(\bar{\theta} + \ell) - \left[\Psi(\bar{\theta} - \hat{\ell}) + \Psi(\bar{\theta} + \hat{\ell})\right] \\
&\Leftrightarrow \int_{\frac{\hat{\ell}}{1-x}}^{\frac{\ell}{1-x}} \left(\int_{-x}^{x} \Psi''\left(\frac{\bar{\theta} - x\eta}{1-x} + y\right) dy\right) dx > \int_{\hat{\ell}}^{\ell} \left(\int_{-x}^{x} \Psi''(\bar{\theta} + y)dy\right) dx. \qquad \text{(A.61)}
\end{aligned}$$

Since $\Psi(\theta)$ is convex and $x \in [0, 1)$,

$$\begin{aligned}
\int_{\frac{\hat{\ell}}{1-x}}^{\frac{\ell}{1-x}} \left(\int_{-x}^{x} \Psi''\left(\frac{\bar{\theta} - x\eta}{1-x} + y\right) dy\right) dx &> \int_{\frac{\hat{\ell}}{1-x}}^{\frac{\hat{\ell}}{1-x}+\ell-\hat{\ell}} \left(\int_{-x}^{x} \Psi''\left(\frac{\bar{\theta} - x\eta}{1-x} + y\right) dy\right) dx \\
&> \int_{\hat{\ell}}^{\ell} \left(\int_{-x}^{x} \Psi''\left(\frac{\bar{\theta} - x\eta}{1-x} + y\right) dy\right) dx.
\end{aligned}$$

Since, in addition,

$$\Psi''(\theta) = \frac{C^2}{4(B + C\theta)^{\frac{3}{2}}} \frac{1}{\left(A + \sqrt{B + C\theta}\right)^2} + \frac{C^2}{2(B + C\theta)} \frac{1}{\left(A + \sqrt{B + C\theta}\right)^3}$$

is decreasing, (A.61) holds under the sufficient condition $\frac{\bar{\theta} - x\eta}{1-x} \leq \bar{\theta}$, which is equivalent to $\bar{\theta} \leq \eta$. $\square$

Proof of Proposition 4.1. We proceed in three steps. In a first step we show that for any $\epsilon > 0$, the investor's expected utility under any feasible contract does not exceed the utility (4.1), achieved when the benchmark position η, the delegation set $\mathcal{L}$ and the manager's positions are as in the proposition, and the fee $f(W)$ is zero. Moreover, the difference in utilities is bounded away from zero when η, $\mathcal{L}$ or the manager's positions are not as in the proposition. In a second step we show that the investor's expected utility under the feasible contract described in the proposition converges to the maximum utility (4.1) when ϵ goes to zero. In a third step we show that the maximum in (4.1) is achieved for $L > 0$.

The first and second steps imply that $(\eta, \mathcal{L})$ described in the proposition are part of an optimal contract when ϵ goes to zero, and that such a contract generates the manager's positions described in the proposition. Indeed, if an optimal contract involved different η, $\mathcal{L}$ or manager's positions, then it would yield a utility bounded away from (4.1), while the feasible contract described in the proposition yields that utility when ϵ goes to zero. Adding the third step implies that the investor employs the manager. Indeed, not employing her is equivalent to setting $(\eta, \mathcal{L}, f(W)) = (z_0^*, \{0\}, 0)$, but this generates a utility bounded away from (4.1) because $\mathcal{L} = \{0\}$ is not as in the proposition.

Step 1: Since the fee $f(W)$ is increasing and Π^O gives positive probability only to positive values of R, an unskilled manager with posterior Π^O chooses the maximum value of z that meets the constraint $F(\Delta, s) \in \mathcal{L}$. Conversely, since $f(W)$ is increasing and Π^P gives positive probability only to negative values of R, an unskilled manager with posterior Π^P chooses the minimum value of z such that $F(\Delta, s) \in \mathcal{L}$. We denote these maximum and minimum values by $\bar{z}(s)$ and $\underline{z}(s)$, respectively, using the same notation as in the case $\mathcal{L} = [0, L]$.

Consider a feasible contract $(f(W), \mathcal{L})$ that induces positions $z(s)$ by the skilled manager (and $\bar{z}(s)$ and $\underline{z}(s)$ by the unskilled manager). Since $f(W)$ is positive (because it is non-negative and increasing), the investor's expected utility is smaller than the utility achieved when the manager's positions remain the same and the fee is zero. The latter utility is

$$-(1-\lambda)\int_{s\in\Phi}\mathbb{E}_s\left(e^{-\rho z(s)SR}\right)h(s)ds - \frac{\lambda}{2}\int_{s\in\Phi}\left[\mathbb{E}_0\left(e^{-\rho\bar{z}(s)SR}\right)+\mathbb{E}_0\left(e^{-\rho\underline{z}(s)SR}\right)\right]h(s)ds. \quad \text{(A.62)}$$

Since the investor's conditional expected utility $-\mathbb{E}_s\left(e^{-\rho zSR}\right)$ is concave in Z, it is increasing for $z < z^*(s)$ and decreasing for $z > z^*(s)$. Therefore, when $z^*(s) > \bar{z}(s)$, expected utility for $z(s)$ is smaller than for $\bar{z}(s) > z(s)$. Conversely, when $z^*(s) < \underline{z}(s)$, expected utility for $z(s)$ is smaller than for $\underline{z}(s) < z(s)$. Since, in addition, expected utility is maximum for $z^*(s)$, (A.62) does not exceed

$$
\begin{aligned}
&-(1-\lambda)\left[\int_{s\in\bar{\Phi}}\mathbb{E}_s\left(e^{-\rho\bar{z}(s)SR}\right)h(s)ds+\int_{s\in\Phi\setminus(\bar{\Phi}\cup\underline{\Phi})}\mathbb{E}_s\left(e^{-\rho z^*(s)SR}\right)h(s)ds\right.\\
&\left.+\int_{s\in\underline{\Phi}}\mathbb{E}_s\left(e^{-\rho\underline{z}(s)SR}\right)h(s)ds\right]-\frac{\lambda}{2}\int_{s\in\Phi}\left[\mathbb{E}_0\left(e^{-\rho\bar{z}(s)SR}\right)+\mathbb{E}_0\left(e^{-\rho\underline{z}(s)SR}\right)\right]h(s)ds, \qquad \text{(A.63)}
\end{aligned}
$$

where we denote by $\bar{\Phi}$ the set of signals s such that $z^*(s)>\bar{z}(s)$, and by $\underline{\Phi}$ the set of signals s such that $z^*(s)<\underline{z}(s)$. Equation (A.63) describes also the expected utility when the set $\mathcal{L}$ is replaced by $[0,L]$ with $L\equiv\sup\mathcal{L}$, since $(\bar{z}(s),\underline{z}(s))$ are the same for both sets. Since replacing $\mathcal{L}$ by $[0,L]$ yields the term in curly brackets in (4.1), and since (4.1) is the maximum of that term over (η,L), it exceeds the utility under any feasible contract.

For (η,L) not maximizing the term in curly brackets in (4.1), (A.63) is smaller than (4.1). For $\mathcal{L}$ differing from $[0,L]$ by a positive-measure set, (A.62) is smaller than (A.63) and hence also than (4.1). Indeed, since the range of $z^*(s)$ is the real line, $z(s)$ differs from $z^*(s)$ in a positive-measure set. For $z(s)$ differing from the values in (A.63) in a positive-measure set (while meeting the constraint $F(\Delta,s)\in\mathcal{L}$), (A.62) is smaller than (A.63) and hence also than (4.1). Since (4.1), (A.62) and (A.63) are independent of ϵ, the difference between (4.1) and the utility under a feasible contract in which η, $\mathcal{L}$ or the manager's positions are not as in the proposition is bounded away from zero.

Step 2: Consider the contract $(f(W),\eta,\mathcal{L})$ with $(\eta,\mathcal{L})$ described in the proposition and

$$
f(W)=\epsilon g(W)+\epsilon^{\frac{1}{2}}\left(\epsilon^{-\frac{1}{8}}-e^{-\rho W}\right)1_{W>\frac{1}{8\rho}\log(\epsilon)}. \qquad \text{(A.64)}
$$

(The term $1_{W>\underline{W}}$ is the indicator function, equal to one if $W>\underline{W}$ and zero otherwise.) Since the function $\epsilon^{-\frac{1}{8}}-e^{-\rho W}$ is positive and increasing for $W>\frac{1}{8\rho}\log(\epsilon)$, the fee $f(W)$ satisfies the non-negativity and monotonicity constraints. Since the function $g(W)$ is bounded over $(-\infty,\infty)$ and the function $1-\epsilon^{\frac{1}{8}}e^{-\rho W}$ is bounded over $W>\frac{1}{8\rho}\log(\epsilon)$, $f(W)$ converges uniformly to zero when ϵ goes to zero.

Equation (A.64) implies that the manager's utility is

$$
-e^{-\bar{\rho}f(W)}=-1+\bar{\rho}\epsilon^{\frac{1}{2}}\left(\epsilon^{-\frac{1}{8}}-e^{-\rho W}\right)1_{W>\frac{1}{8\rho}\log(\epsilon)}+\epsilon^{\frac{3}{4}}k(W), \qquad \text{(A.65)}
$$

where the function $k(W)$ is uniformly bounded when ϵ goes to zero. Since the dominant term in (A.65) in the interval $W>\frac{1}{4\rho}\log(\epsilon)$ is an affine transformation of the investor's utility, the position that maximizes the skilled manager's expected utility when ϵ goes to zero converges to the position that maximizes the investor's expected utility. Hence, when ϵ goes to zero, the investor's expected utility is given by the (4.1).

Step 3: Breaking the integrals, we can write the term in curly brackets in (4.1) as

$$\begin{aligned}&-(1-\lambda)\left[\int_{s\in\bar{\Phi}}\mathbb{E}_s\left(e^{-\rho\bar{z}(s)SR}\right)h(s)ds+\int_{s\in\Phi\backslash(\bar{\Phi}\cup\underline{\Phi})}\mathbb{E}_s\left(e^{-\rho z^*(s)SR}\right)h(s)ds\right.\\&\left.+\int_{s\in\underline{\Phi}}\mathbb{E}_s\left(e^{-\rho\underline{z}(s)SR}\right)h(s)ds\right]-\frac{\lambda}{2}\int_{s\in\Phi}\left[\mathbb{E}_0\left(e^{-\rho\bar{z}(s)SR}\right)+\mathbb{E}_0\left(e^{-\rho\underline{z}(s)SR}\right)\right]h(s)ds.\end{aligned}\tag{A.66}$$

The derivative of (A.66) with respect to L is

$$\begin{aligned}&(1-\lambda)\left[\int_{s\in\bar{\Phi}}\mathbb{E}_s\left(e^{-\rho\bar{z}(s)SR}R\right)S\frac{\partial\bar{z}(s)}{\partial L}h(s)ds+\int_{s\in\underline{\Phi}}\mathbb{E}_s\left(e^{-\rho\underline{z}(s)SR}R\right)S\frac{\partial\underline{z}(s)}{\partial L}h(s)ds\right]\\&+\frac{\lambda}{2}\int_{s\in\Phi}\left[\mathbb{E}_0\left(e^{-\rho\bar{z}(s)SR}SR\frac{\partial\bar{z}(s)}{\partial L}\right)+\mathbb{E}_0\left(e^{-\rho\underline{z}(s)SR}SR\frac{\partial\underline{z}(s)}{\partial L}\right)\right]h(s)ds.\end{aligned}\tag{A.67}$$

The definitions of $(\bar{z}(s),\underline{z}(s))$ imply

$$\frac{\partial\bar{z}(s)}{\partial L}=-\frac{\partial\underline{z}(s)}{\partial L}>0.\tag{A.68}$$

Together with the definitions of $(\bar{\Phi},\underline{\Phi})$, they also imply $\mathbb{E}_s\left(e^{-\rho\bar{z}(s)SR}R\right)>0$ for all $s\in\bar{\Phi}(L)$, and $\mathbb{E}_s\left(e^{-\rho\underline{z}(s)SR}R\right)<0$ for all $s\in\underline{\Phi}(L)$. Hence, the first of the two terms in square brackets in (A.67) is positive. The second term is zero for $L=0$ because of (A.68) and $\bar{z}(s)=\underline{z}(s)=\eta$. Therefore, the derivative of (A.66) with respect to L is positive at $L=0$, which means that the maximum in (4.1) is achieved for $L>0$. $\square$

Proof of Proposition 4.2. The maximum position $\bar{z}(s)$ and minimum position $\underline{z}(s)$ that meet the constraint

$$F(\Delta,s)=|z_t-\eta|\sigma\sqrt{D_t}S'(D_t)\leq L$$

are

$$\bar{z}(s)=\eta+\frac{L}{\sigma\sqrt{D_t}S'(D_t)},\tag{A.69}$$

$$\underline{z}(s)=\eta-\frac{L}{\sigma\sqrt{D_t}S'(D_t)},\tag{A.70}$$

respectively. The position $z^*(s)$ that maximizes an investor's expected utility conditional on s is the position that unconstrained investors hold in equilibrium. In the unconstrained region, defined by (4.3), $z^*(s)$ can be derived by setting $z_{1t} = z_{2t}$ in the market-clearing condition (4.2), and is

$$z_{1t} = z^*(s) = \frac{\theta - \lambda x\eta}{1 - \lambda x}.$$

To derive $z^*(s)$ in the constrained region, defined by

$$\frac{|\theta - \eta|}{1 - \lambda x}\sigma\sqrt{D_t}S'(D_t) > L,$$

we distinguish cases. When $\theta > \eta$, $z^*(s)$ can be derived by setting $z_{2t} = \bar{z}(s)$ in (4.2), and is

$$z_{1t} = z^*(s) = \frac{\theta - x\eta - \frac{(1-\lambda)xL}{\sigma\sqrt{D_t}S'(D_t)}}{1 - x}.$$

When instead $\theta < \eta$, $z^*(s)$ can be derived by setting $z_{2t} = \underline{z}(s)$ in (4.2), and is

$$z_{1t} = z^*(s) = \frac{\theta - x\eta + \frac{(1-\lambda)xL}{\sigma\sqrt{D_t}S'(D_t)}}{1 - x}.$$

Hence, $(\bar{z}(s), \underline{z}(s), z^*(s))$ are as in the proposition's statement. The ODEs in the unconstrained and constrained region can be derived from (3.2) by substituting $\mathbb{E}_t(dR_t^{sh})$ by the drift term in (3.1), $\mathbb{V}\mathrm{ar}_t(dR_t^{sh})$ by the square of the diffusion term, and z_{1t} by $z^*(s)$. This yields

$$D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t) = \frac{\rho(\theta - \lambda x\eta)}{1 - \lambda x}\sigma^2 D_t S'(D_t)^2 \tag{A.71}$$

in the unconstrained region, and

$$D_t + \kappa(\bar{D} - D_t)S'(D_t) + \frac{1}{2}\sigma^2 D_t S''(D_t) - rS(D_t) = \frac{\rho(\theta - x\eta)}{1 - x}\sigma^2 D_t S'(D_t)^2 - \frac{\rho\,\mathrm{sgn}(\theta - \eta)(1 - \lambda)xL}{1 - x}\sigma\sqrt{D_t}S'(D_t) \tag{A.72}$$

in the constrained region. When $\theta > \eta$, $z^*(s) \in (\eta, \bar{z}(s)]$ in the unconstrained region, and $z^*(s) > \bar{z}(s)$ in the constrained region. When instead $\theta < \eta$, $z^*(s) \in [\underline{z}(s), \eta)$ in the unconstrained region, and $z^*(s) < \underline{z}(s)$ in the constrained region. Hence,

$$\bar{\Phi} = \{\theta > \eta \text{ and } z^*(s) > \bar{z}(s)\} = \{\theta > \eta \text{ and } \frac{|\theta - \eta|}{1 - \lambda x}\sigma\sqrt{D_t}S'(D_t) > L\}, \tag{A.73}$$

$$\underline{\Phi} = \{\theta < \eta \text{ and } z^*(s) < \bar{z}(s)\} = \{\theta < \eta \text{ and } \frac{|\theta - \eta|}{1 - \lambda x}\sigma\sqrt{D_t}S'(D_t) > L\}. \tag{A.74}$$

The derivative of (A.66) with respect to L in the continuous-time limit can be derived from

(A.67) by replacing SR by dR_t^{sh}, and is

$$(1-\lambda)\left[\int_{s\in\bar{\Phi}}\mathbb{E}_s\left(e^{-\rho\bar{z}(s)dR_t^{sh}}dR_t^{sh}\right)\frac{\partial\bar{z}(s)}{\partial L}h(s)ds+\int_{s\in\underline{\Phi}}\mathbb{E}_s\left(e^{-\rho\underline{z}(s)dR_t^{sh}}dR_t^{sh}\right)\frac{\partial\underline{z}(s)}{\partial L}h(s)ds\right]$$
$$+\frac{\lambda}{2}\int_{s\in\Phi}\left[\mathbb{E}_0\left(e^{-\rho\bar{z}(s)dR_t^{sh}}dR_t^{sh}\frac{\partial\bar{z}(s)}{\partial L}\right)+\mathbb{E}_0\left(e^{-\rho\underline{z}(s)dR_t^{sh}}dR_t^{sh}\frac{\partial\underline{z}(s)}{\partial L}\right)\right]h(s)ds. \tag{A.75}$$

The derivative of (A.66) with respect to η in the continuous-time limit is

$$(1-\lambda)\left[\int_{s\in\bar{\Phi}}\mathbb{E}_s\left(e^{-\rho\bar{z}(s)dR_t^{sh}}dR_t^{sh}\right)\frac{\partial\bar{z}(s)}{\partial \eta}h(s)ds+\int_{s\in\underline{\Phi}}\mathbb{E}_s\left(e^{-\rho\underline{z}(s)dR_t^{sh}}dR_t^{sh}\right)\frac{\partial\underline{z}(s)}{\partial \eta}h(s)ds\right]$$
$$+\frac{\lambda}{2}\int_{s\in\Phi}\left[\mathbb{E}_0\left(e^{-\rho\bar{z}(s)dR_t^{sh}}dR_t^{sh}\frac{\partial\bar{z}(s)}{\partial \eta}\right)+\mathbb{E}_0\left(e^{-\rho\underline{z}(s)dR_t^{sh}}dR_t^{sh}\frac{\partial\underline{z}(s)}{\partial \eta}\right)\right]h(s)ds. \tag{A.76}$$

To simplify (A.75) and (A.76), we use

$$\begin{aligned}\mathbb{E}_s\left(e^{-\rho z(s)dR_t^{sh}}dR_t^{sh}\right)&=\mathbb{E}_s\left((1-\rho z(s)dR_t^{sh})dR_t^{sh}\right)\\&=\mathbb{E}_s(dR_t^{sh})-\rho z(s)\mathbb{V}\mathrm{ar}_s(dR_t^{sh})\\&=\rho[z^*(s)-z(s)]\mathbb{V}\mathrm{ar}_s(dR_t^{sh})\\&=\rho[z^*(s)-z(s)]\sigma^2D_tS'(D_t)^2,\end{aligned} \tag{A.77}$$

where the third step follows because $z^*(s)$ is optimal and hence satisfies the first-order condition (3.2);

$$\begin{aligned}\mathbb{E}_0\left(e^{-\rho z(s)dR_t^{sh}}dR_t^{sh}F(s)\right)&=\mathbb{E}_0\left[\mathbb{E}_s\left(e^{-\rho z(s)dR_t^{sh}}dR_t^{sh}F(s)\right)\right]\\&=\rho\mathbb{E}_0\left[\mathbb{E}_s\left([z^*(s)-z(s)]\sigma^2D_tS'(D_t)^2F(s)\right)\right]\\&=\rho\mathbb{E}_0\left[[z^*(s)-z(s)]\sigma^2D_tS'(D_t)^2F(s)\right],\end{aligned} \tag{A.78}$$

where the second step follows from (A.77); and

$$\frac{\partial\bar{z}(s)}{\partial L}=\frac{1}{\sigma\sqrt{D_t}S'(D_t)}, \tag{A.79}$$
$$\frac{\partial\underline{z}(s)}{\partial L}=-\frac{1}{\sigma\sqrt{D_t}S'(D_t)}, \tag{A.80}$$
$$\frac{\partial\bar{z}(s)}{\partial \eta}=1, \tag{A.81}$$
$$\frac{\partial\underline{z}(s)}{\partial \eta}=1, \tag{A.82}$$

which follow by differentiating (A.69) and (A.70).

Using (A.73), (A.74) and (A.77)-(A.80), we can write (A.75) as

$$\sum_{\theta>\eta}\pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\bar{z}(s)]1_{\{z^*(s)>\bar{z}(s)\}}\sigma\sqrt{D_t}S'(D_t)\right]-\sum_{\theta<\eta}\pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\underline{z}(s)]1_{\{z^*(s)<\underline{z}(s)\}}\sigma\sqrt{D_t}S'(D_t)\right]$$
$$+\frac{\lambda}{2(1-\lambda)}\sum_{\theta}\pi(\theta)\mathbb{E}_\theta\left[[\underline{z}(s)-\bar{z}(s)]\sigma\sqrt{D_t}S'(D_t)\right]. \tag{A.83}$$

Setting (A.83) to zero, and using (A.69) and (A.70) to simplify the third term, we find (4.5). Using (A.73), (A.74), (A.77), (A.78), (A.81) and (A.82), we can write (A.76) as

$$\sum_{\theta>\eta}\pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\bar{z}(s)]1_{\{z^*(s)>\bar{z}(s)\}}\sigma^2D_tS'(D_t)^2\right]+\sum_{\theta<\eta}\pi(\theta)\mathbb{E}_\theta\left[[z^*(s)-\underline{z}(s)]1_{\{z^*(s)<\underline{z}(s)\}}\sigma^2D_tS'(D_t)^2\right]$$
$$+\frac{\lambda}{2(1-\lambda)}\sum_{\theta}\pi(\theta)\mathbb{E}_\theta\left[[2z^*(s)-\bar{z}(s)-\underline{z}(s)]\sigma^2D_tS'(D_t)^2\right]. \tag{A.84}$$

Setting (A.84) to zero, and using (A.69) and (A.70) to simplify the third term, we find (4.4).

When $\eta>\theta_{max}$, the first term in the left-hand side of (4.4) is zero because the summation is over an empty set of θ, the second term is negative because the summation is over a non-empty set of θ and the set of values of D_t such that $z^*(s)<\underline{z}(s)$ has positive measure, and the third term is negative because $z^*(s)<\eta$ when $\theta<\eta$. Hence, the left-hand side of (4.4) is negative, which means that the investor can raise his utility by lowering η. When instead $\eta<\theta_{min}$, the first term is positive because the summation is over an non-empty set of θ and the set of values of D_t such that $z^*(s)>\bar{z}(s)$ has positive measure, the second term is zero because the summation is over an empty set of θ, and the third term is positive because $z^*(s)>\eta$ when $\theta>\eta$. Hence, the left-hand side of (4.4) is positive, which means that the investor can raise his utility by raising η. Therefore, $\eta\in[\theta_{min},\theta_{max}]$.

When θ can take only one value, θ_{min} and θ_{max} coincide with that value, and so does $\eta\in[\theta_{min},\theta_{max}]$. Moreover, the first and second terms in the left-hand side of (4.5) are zero because the summations are over empty sets of θ. Hence, $L=0$.

When θ can take multiple values, the argument showing that the left-hand side of (4.4) is negative when $\eta>\theta_{max}$ can be extended to $\eta\geq\theta_{max}$ because the set of $\theta<\eta$ is non-empty. Likewise, the argument showing that the left-hand side of (4.4) is positive when $\eta<\theta_{min}$ can be extended to $\eta\leq\theta_{min}$ because the set $\theta>\eta$ is non-empty. Therefore, $\eta\in(\theta_{min},\theta_{max})$. Fixing $\eta\in(\theta_{min},\theta_{max})$, the first and second terms in the left-hand side of (4.5) are positive and bounded for $L\geq 0$, and converge to zero when L goes to infinity. When $\lambda=0$, the third term is zero. Hence, the left-hand side of (4.5) is positive, which means that the investor can raise his utility by raising L to infinity. When $\lambda\in[0,1)$, the third term is a linear and decreasing function of L. Hence, the solution L to (4.5) is finite. When λ goes to one, the third term converges to infinity for any finite L. Hence, the solution L to (4.5) converges to zero. $\square$

B Alternative Constraint

We first show that (2.7) resembles closely a constraint restricting the weights of a portfolio consisting of one riskless and multiple risky assets. Suppose that there are N risky assets instead of one. Denote by S_{nt} the price of risky asset $n = 1, .., N$, by z_{nt} the number of shares of the asset that an investor is holding, and by η_n the number of shares of the asset in the benchmark portfolio. Denote also by η_0 the riskless asset investment in the benchmark portfolio. The weight of risky asset n in the investor's portfolio is $\frac{z_{nt}S_{nt}}{W_t}$ and in the benchmark portfolio is $\frac{\eta_{nt}S_{nt}}{\eta_0+\sum_{n=1}^{N}\eta_{nt}S_{nt}}$, where W_t is the investor's wealth. Restricting the two weights not to differ by more than $L \geq 0$ yields the constraint

$$\left|\frac{z_{nt}S_{nt}}{W_t} - \frac{\eta_n S_{nt}}{\eta_0 + \sum_{n=1}^{N}\eta_n S_{nt}}\right| \leq L \Leftrightarrow \left|z_{nt} - \frac{\eta_n W_t}{\eta_0 + \sum_{n=1}^{N}\eta_n S_{nt}}\right| S_{nt} \leq LW_t. \tag{B.1}$$

The constraint (2.7) is the same as (B.1) except that η_n and L replace $\frac{\eta_n W_t}{\eta_0+\sum_{n=1}^{N}\eta_n S_{nt}}$ and LW_t, respectively. The quantity $\frac{\eta_n W_t}{\eta_0+\sum_{n=1}^{N}\eta_n S_{nt}}$ is the number of shares of asset n that the investor holds when investing all his wealth W_t in the benchmark portfolio. The constraint (2.7) approximates (B.1) well when the relative variation in $\frac{W_t}{\eta_0+\sum_{n=1}^{N}\eta_n S_{nt}}$ and W_t is small compared to that in S_{nt}. That condition is likely to hold when N is sufficiently large.

In a similar spirit, (2.6) resembles closely a constraint restricting the volatility of a portfolio consisting of one riskless and multiple risky assets. Restricting the volatility of each component of the portfolio separately yields the constraint

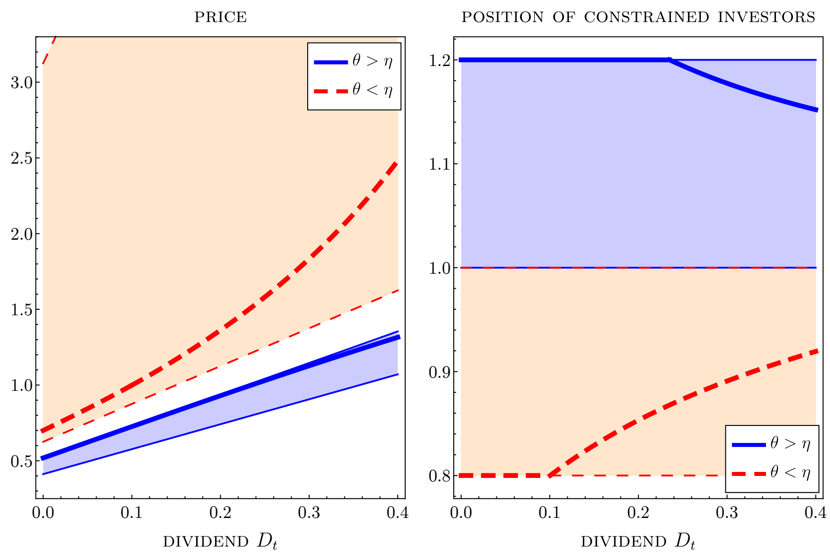
$$\left|\frac{z_{nt}S_{nt}}{W_t} - \frac{\eta_n S_{nt}}{\eta_0 + \sum_{n=1}^{N}\eta_n S_{nt}}\right| \sqrt{\frac{\mathbb{V}\mathrm{ar}_t(R_{nt})}{dt}} \leq L \Leftrightarrow \left|z_{nt} - \frac{\eta_n W_t}{\eta_0 + \sum_{n=1}^{N}\eta_n S_{nt}}\right| \sqrt{\frac{\mathbb{V}\mathrm{ar}_t(R^{sh}_{nt})}{dt}} \leq LW_t. \tag{B.2}$$

The constraint (2.6) is the same as (B.2) except that η_n and L replace $\frac{\eta_n W_t}{\eta_0+\sum_{n=1}^{N}\eta_n S_{nt}}$ and LW_t, respectively.

Under the alternative constraint (2.7), the system of ODEs consists of (3.4) in the unconstrained region defined by

$$|\theta - \eta| S(D_t) \leq L, \tag{B.3}$$

Figure 6: Alternative Constraint



Asset price S_t (left panel) and position z_{2t} of constrained investors (right panel) as functions of the dividend flow D_t, under the alternative constraint (2.7). The thick lines correspond to the case where there is a constraint ($L \in (0, \infty)$). The thin lines correspond to the polar cases where there is no constraint ($L = \infty$) and where the constraint is infinitely tight ($L = 0$). The latter case corresponds to the more extreme values in the left panel and to the horizontal line $z_{2t} = 1$ in the right panel. In all three cases, the dashed red line is drawn for $\theta = 0.8$ and the solid blue line is drawn for $\theta = 1.2$. All parameter values are as in Figure 1 except for L which is equal to 0.2.

and

$$D_t+\kappa(\bar{D}-D_t)S'(D_t)+\frac{1}{2}\sigma^2 D_t S''(D_t)-rS(D_t) = \frac{\rho(\theta-x\eta)}{1-x}\sigma^2 D_t S'(D_t)^2-\frac{\rho\,\mathrm{sgn}(\theta-\eta)xL}{1-x}\frac{\sigma^2 D_t S'(D_t)^2}{S(D_t)} \tag{B.4}$$

in the constrained region defined by

$$|\theta-\eta|S(D_t) > L. \tag{B.5}$$

Figure 6 plots the solution of that system within the numerical example of Section 3. The left panel plots the price and the right panel plots the position of constrained investors, both as function of D_t. Consistent with Theorem 3.1, the price for $L \in (0, \infty)$ lies between the affine solution in the polar cases where there is no constraint ($L = \infty$) and where the constraint is infinitely tight ($L = 0$). Moreover, the price is convex in D_t for $\theta < \eta$ and concave for $\theta > \eta$.

References

Admati, Anat, and Paul Pfleiderer, 1997, Does it all add up? Benchmarks and the compensation of active portfolio managers, *Journal of Business* 70, 323–350.

Alonso, Ricardo, and Niko Matouschek, 2008, Optimal delegation, *Review of Economic Studies* 75, 259–293.

Amador, Manuel, and Kyle Bagwell, 2013, The theory of optimal delegation with an application to tariff caps, *Econometrica* 81, 1541–1599.

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler, 2011, Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly, *Financial Analysts Journal* 67.

Basak, Suleyman, and Anna Pavlova, 2013, Asset prices and institutional investors, *American Economic Review* 103, 1728–1758.

Berk, Jonathan, and Richard Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Bhattacharya, Ayan, and Maureen O'Hara, 2018, Can ETFs increase market fragility? Effects of information linkages in ETF markets, Working paper Cornell University.

Bhattacharya, Sudipto, and Paul Pfleiderer, 1985, Delegated portfolio management, *Journal of Economic Theory* 36, 1–25.

BIS, 2003, Incentive structures in institutional asset management and their implications for financial markets, Report of the Committee on the Global Financial System.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–55.

———, Michael Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael Jensen, ed.: *Studies in the Theory of Capital Markets* . pp. 79–121 (Praeger: New York).

Bond, Philip, and Diego Garcia, 2019, The equilibrium consequences of indexing, Working paper University of Washington.

Brennan, Michael, 1993, Agency and asset pricing, Working paper 1147, Anderson Graduate School of Management, UCLA.

Brunnermeier, Markus, and Yuliy Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104, 379–421.

Buffa, Andrea, and Idan Hodor, 2018, Institutional investors, heterogeneous benchmarks and the comovement of asset prices, Working paper Boston University.

Buffa, Andrea, Dimitri Vayanos, and Paul Woolley, 2014, Asset management contracts and equilibrium prices, Working paper London School of Economics.

Cadenillas, Abel, Jaksa Cvitanic, and Fernando Zapatero, 2007, Optimal risk-sharing with effort and project choice, *Journal of Economic Theory* 133, 403–440.

Christoffersen, Susan, and Mikhail Simutin, 2017, On the demand for high-beta stocks: Evidence from mutual funds, *Review of Financial Studies* 30, 2596–2620.

Cong, William, and Douglas Xu, 2016, Rise of factor investing: Asset prices, informational efficiency and security design, Working paper Cornell University.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 475–512.

Cuoco, Domenico, and Ron Kaniel, 2011, Equilibrium prices in the presence of delegated portfolio management, *Journal of Financial Economics* 101, 264–296.

Cvitanic, Jaksa, and Hao Xing, 2018, Asset pricing under optimal contracts, *Journal of Economic Theory* 173, 142–180.

Das, Sanjiv, and Rangarajan Sundaram, 2002, Fee speech: Signaling, risk-sharing, and the impact of fee structures on investor welfare, *Review of Financial Studies* 15, 1465–1497.

Dasgupta, Amil, and Andrea Prat, 2008, Information aggregation in financial markets with career concerns, *Journal of Economic Theory* 143, 83–113.

———, and Michela Verardo, 2011, The price impact of institutional herding, *Review of Financial Studies* 24, 892–925.

Dybvig, Philip, Heber Farnsworth, and Jennifer Carpenter, 2010, Portfolio performance and agency, *Review of Financial Studies* 23, 1–23.

Elton, Edwin, and Martin Gruber, 2013, Mutual funds, in *Handbook of the Economics of Finance* . chap. 15, pp. 1011–1061 (Elsevier: London/New York).

Franzoni, Francesco, Itzhak Ben-David, and Rabih Moussawi, 2017, Exchange-traded funds, *Annual Review of Financial Economics* 9, 169–189.

Frazzini, Andrea, and Owen Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

French, Kenneth, 2008, Presidential address: The cost of active investing, *Journal of Finance* 63, 1537–1573.

Froot, Kenneth, David Scharftstein, and Jeremy Stein, 1992, Herd on the street: Informational inefficiencies in a market with short-term speculation, *Journal of Finance* 47, 1461–84.

Garcia, Diego, and Joel Vanden, 2009, Information acquisition and mutual funds, *Journal of Economic Theory* 144, 1965–1995.

Garleanu, Nicolae, and Lasse Pedersen, 2018, Efficiently inefficient markets for assets and asset management, *Journal of Finance* 73, 1663–1712.

Gorton, Gary, Ping He, and Lixin Huang, 2010, Security price informativeness with delegated traders, *American Economic Journal: Microeconomics* 2, 137–170.

Gromb, Denis, and Dimitri Vayanos, 2010, Limits of arbitrage, *Annual Review of Financial Economics* 2, 251–275.

Grossman, Sanford, and Joseph Stiglitz, 1980, On the impossibility of informationnaly efficient markets, *American Economic Review* 70, 393–408.

Guerrieri, Veronica, and Peter Kondor, 2012, Fund managers, career concerns, and asset price volatility, *American Economic Review* 102, 1986–2017.

Harrison, Michael, and David Kreps, 1978, Speculative investor behavior in a stock market with heterogeneous expectations, *Quarterly Journal of Economics* 92, 323–336.

Haugen, Robert, and Nardin Baker, 1996, Commonality in the determinant of expected stock returns, *Journal of Financial Economics* 41, 401–439.

He, Zhiguo, and Arvind Krishnamurthy, 2012, A model of capital and crises, *Review of Economic Studies* 79, 735–777.

———, 2013, Intermediary asset pricing, *American Economic Review* 103, 1–43.

He, Zhiguo, and Wei Xiong, 2013, Delegated asset management, investment mandates, and capital immobility, *Journal of Financial Economics* 107, 239–258.

Hong, Harrison, and Marcin Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hong, Harrison, and Jeremy Stein, 2007, Disagreement and the stock market, *Journal of Economic Perspectives* 21, 109–128.

Huang, Shiyang, 2018, Delegated information acquisition and asset pricing, Working paper Hong Kong University.

Jones, Bradley, 2015, Asset bubbles: Re-thinking policy for the age of asset management, Working paper International Monetary Fund.

Kapur, Sandeep, and Allan Timmermann, 2005, Relative performance evaluation contracts and asset market equilibrium, *Economic Journal* 115, 1077–1102.

Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova, 2018, The benchmark inclusion subsidy, Working paper University of Chicago.

Kyle, Albert, Hui Ou-Yang, and Bin Wei, 2011, A model of portfolio delegation and strategic trading, *Review of Financial Studies* 24, 3778–3812.

Li, Wei, and Ashish Tiwari, 2009, Incentive contracts in delegated portfolio management, *Review of Financial Studies* 22, 4681–4714.

Lines, Anton, 2016, Do institutional incentives distort asset prices?, Working paper Columbia University.

Malamud, Semyon, and Evgeny Petrov, 2014, Portfolio delegation and market efficiency, Working paper Swiss Finance Institute.

Ou-Yang, Hui, 2003, Optimal contracts in a continuous-time delegated portfolio management problem, *Review of Financial Studies* 16, 173–208.

Palomino, Frederic, and Andrea Prat, 2003, Risk taking and optimal contracts for money managers, *RAND Journal of Economics* 34, 113–137.

Pastor, Lubos, and Robert Stambaugh, 2012, On the size of the active management industry, *Journal of Political Economy* 120, 740–781.

Qiu, Zhigang, 2017, Equilibrium-informed trading with relative performance measurement, *Journal of Financial and Quantitative Analysis* 52, 2083–2118.

Sato, Yuki, 2016, Delegated portfolio management, optimal fee contracts, and asset prices, *Journal of Economic Theory* 165, 360–389.

Scheinkman, Jose, and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1219.

Sockin, Michael, and Mindy Zhang Xiaolan, 2018, Delegated learning in asset management, Working paper University of Texas at Austin.

Stambaugh, Robert, 2014, Presidential address: Investment noise and trends, *Journal of Finance* 69, 1415–1453.

Starks, Laura, 1987, Performance incentive fees: An agency theoretic approach, *Journal of Financial and Quantitative Analysis* 22, 17–32.

Stoughton, Neal, 1993, Moral hazard and the portfolio management problem, *Journal of Finance* 48, 2009–2028.

Subrahmanyam, Avanidhar, 1991, A theory of trading in stock index futures, *Review of Financial Studies* 4, 17–51.

Vayanos, Dimitri, 2018, Risk limits as optimal contracts, Working paper London School of Economics.