# High Water Marks in Competitive Capital Markets

Susan E. K. Christoffersen

David K. Musto

Bilge Yilmaz*

December 2, 2015

**ABSTRACT**

We analyze the effectiveness of hedge funds' high-water mark provisions (HWMs) at mitigating moral hazard and adverse selection, when the supply of capital is competitive. We show that the HWM does not allow better managers to distinguish themselves ex ante. Regarding moral hazard, we find that precision of beliefs about a manager's ability is crucial to the HWM's effectiveness. When precision is low, the HWM encourages shirking. However, if precision is high, and the beliefs themselves are high as well, then the HWM encourages costly effort by the manager. Furthermore, the HWM boosts the initial fund size, depresses initial expected returns and increases subsequent expected returns.

*Christoffersen is from the Rotman School of Management and Copenhagen Business School. Musto and Yilmaz are from the Wharton School at the University of Pennsylvania. We are grateful for comments from Rick Green, David Hirschleifer, Sugata Ray, Fabrice Rouah, Luke Taylor and seminar participants at Beijing University, Boston University, Cambridge, Carnegie Mellon, Cheung Kong GSB China, Copenhagen Business School, Cornell, Georgetown, European Finance Meetings 2009, Georgia State University, Hong Kong Polytechnic, New York University, Kepos Capital Management, Penn Law, Stockholm School of Economics, UBC Summer Conference 2011, UC Irvine, University of Wisconsin at Madison, Vanderbilt, Wharton School, and Wilfred Laurier University.

For an investment in a hedge fund, the investor typically pays both a base fee and an incentive fee, where the incentive fee is subject to a high-water mark provision (HWM). The job of the HWM is to apply this fee only to net profits, i.e. to suspend it until past losses are made up. Because hedge funds enjoy wide freedom in choosing their fee structures, and because they choose HWMs so frequently, it would seem that the HWM brings a benefit. In this paper we address the economic role of the HWM, deriving its implications for assets and expected returns, and also its effect on contract efficiency. Our analysis finds significant implications for assets and expected returns in both the time series and the cross section, while casting doubt on the HWM's ability to improve efficiency when a manager's abilities are not already well known.

Bad fund performance is central to the analysis, as it not only triggers the HWM, but also delivers two other simultaneous effects in opposite directions. One effect is through learning: to the extent that investors learn about a manager from his performance, bad performance reduces his fund's prospects. The other effect is through scale: to the extent that a fund's size dilutes value added per dollar, as explored by Berk and Green (2004), bad performance improves its prospects. Accordingly, our analysis allows for all these effects to play out: bad performance triggers the HWM, which influences subsequent investment, which in turn influences the expected return.

Another key element of the analysis is the HWM's disparate treatment of the investors within a hedge fund. Among mutual funds, by the Berk and Green (2004) analysis, investors equalize prospects with their reactions to past returns. This equalization hinges on a fund's entering investments being treated the same as its incumbent investments: when the expected return of new money reaches zero, the expected return of incumbent money reaches zero too. But a HWM disrupts this logic, because while an investment that makes a loss pays no incentive fee unless and until that loss is made up, any new investment has no losses to make up, so its expected incentive fee is higher, and expected return is lower. So when new money after a loss drives its expected return to zero, the expected return on incumbent investment is positive. To a prospective investor, this means the HWM creates the possibility of a future above-market expected return that can come only after a negative return. It is this possibility that we find to be key to the HWM's effect on the cross section and time series of returns, on fund size and on the manager's effort.

For some intuition on size and expected returns, consider managers $A$ and $B$, who have the same expected value-added, i.e. alpha, but $A$'s alpha is known with complete precision, whereas there is considerable uncertainty about $B$'s. So the effect of a bad return on $A$'s fund is just the positive scale effect, whereas the effect on $B$ includes both the positive scale and the negative learning effect. Suppose this negative effect is strong enough that incumbent investors' expected returns are negative, even with the HWM and the scale effect, after a bad return. Then $B$'s investors will have to withdraw funds just to bring their expectations to zero, through the scale effect. New money would not enter, since its expectation would be lower, and therefore negative. By contrast, $A$'s fund will attract new investment after bad performance, as it is a depleted fund with unchanged prospects, so new investors refill it until their expected return hits zero. Because incumbent investors pay no incentive fee while under the high-water mark, the return expectations of incumbent money are above zero even after the entrance of new money. If the initial return had instead been positive, this possibility of a positive expected return would not exist for any investors, because without losses, new and incumbent money have the same HWM, so new money drives all expectations to zero. Backing up to the original investments with $A$ and $B$, only the investors with $A$ place positive probability on a positive future expected return, so only A's investors compete their initial expected return below zero through additional investment. Thus, the HWM alters expected returns only when investors are relatively certain of the manager's ability, since that is when it delivers a positive expectation after bad performance and shelters this expectation from dilution by new money.

Regarding efficiency, we ask whether the HWM helps solve two standard money-management agency problems: adverse selection and moral hazard. Managers privately know, in the model, whether they are skilled or unskilled, where the skilled add more value, and only if they make a privately costly effort. Thus, we can address whether the HWM reduces adverse selection by screening the unskilled managers out. We find it does not. We can also address whether the HWM reduces moral hazard by increasing the maximum effort cost that the skilled manager will pay. We find that the HWM can increase this cost, but only when the incidence of skilled managers is high, maybe implausibly high.

We can get some intuition for the moral hazard result from the example with managers $A$ and $B$. Because his investors' expected return after a bad performance is zero, it follows that the manager with the imprecisely-known alpha, manager $B$, gets all the expected surplus after bad performance. But whereas this compensation would arrive through both a base and an incentive fee without a HWM, with the HWM it arrives through just the base fee. So the HWM lowers expected incentive pay without touching expected compensation, thereby reducing the reward to effort after a bad return. After a good return it has no effect, so the net effect is simply negative. By contrast, the HWM gives $A$'s investors positive expected returns after bad performance, thereby reducing the manager's expected surplus after bad performance. This punishment raises the possibility that the HWM increases the reward to effort by such managers, who in practice could be the managers with longer track records. We find that this can indeed happen, but it is potentially important that the HWM also reduces the reward for a good initial return, due to the larger base fee that arises from the larger initial investment, and that reduces the net profits the incentive fee is applied to.

Our analysis of the efficiency of the HWM contributes to a literature that to date has explored hedge-fund contracts through the analysis of an initial investment which may eventually leave. This literature has focused particular attention on the manager's incentive to add risk. Standard option-pricing results suggest that the base/incentive fee structure encourages risk taking, but a series of papers analyzing the risk choice from different perspectives (Hodder and Jackwerth, 2007, Panageas and Westerfield, 2008, Chakraborty and Ray, 2010, Drechsler, 2011) shows this intuition to be fragile, particularly when the analysis includes the HWM. Similarly, Carpenter (2000), Ross (2004) and Basak, Pavlova and Shapiro (2007) demonstrate the distortive effect of feeding convex compensation through concave utility. A seminal paper in this area is Goetzmann, Ingersoll and Ross (2003), which focuses primarily on the valuation of the fees arising from a hedge-fund investment.

The analysis most related to ours may be Aragon and Qian (2010), which asks from another perspective how the HWM can improve efficiency. If investors can learn from and leave after a period of performance, the HWM, they show, discourages costly liquidation by bringing the expected fee in line with the fund's reduced prospects.

The rest of the paper is as follows. Section 1 presents the model and some preliminary results. Section 2 addresses the efficiency of the HWM and Section 3 relates the model to empirical findings as well as discusses the regions when the HWM helps to improve moral hazard problems. Finally, Section 4 summarizes and concludes.

# I. Model

## A. The Setup

There are three dates, 0, 1 and 2, with the first period running from 0 to 1 and the second from 1 to 2. A money manager can manage in both periods, and in each period his dollar return is either $\sigma$ or $-\sigma$, which we refer to as a high or low return, respectively.[1]

[1] We capture decreasing returns to scale by assuming that the manager adds a fixed amount of expected value, so that extra investment spreads this expectation across more dollars. See Berk and

The manager has no money, and there are arbitrarily many competitive investors who have money they can invest with the manager at both dates 0 and 1. Everybody is risk-neutral,[2] discounts at zero and has a reservation utility and expected return of zero.

The manager could be a good type ($G$) or a bad type ($B$). Let $\lambda \in (0, 1)$ stand for the probability that the manager is type $G$. A manager knows his type but investors share this common prior. A type $G$ can invest in his trading strategy and improve his chances of success in both periods at a cost $c > 0$ at date 0 whereas a type $B$ cannot improve his chances of success. The manager will have a high return with probability $\theta \in (0, 1)$ if he is type $G$ and makes the costly effort. Otherwise, the probability of high return is $\phi < \theta$, where $\phi \in (0, 1)$.[3] We assume that it is efficient to invest in the trading strategy, i.e. to make the costly effort, which amounts to assuming that $\min\{4(\theta - \phi)\sigma, 2(2\theta - 1)\sigma\} > c$. The first part of the inequality ensures that the expected payoff difference between making and not making the costly effort is higher than the cost of effort. The second inequality ensures that the expected payoff of the good type is positive net of the costly effort.

We denote the amount invested with the manager at date $t$ as $I_t$. The money management contract is exogenously specified as a base fee of $x > 0$ paid at the beginning of a period and an incentive fee $y > 0$ of net profits paid at the end of the period, where the incentive fee may have a HWM. Thus at date 0 the manager gets a base fee of $xI_0$, and at date 1 the manager gets $y(\sigma - xI_0)$ if the first-period return was high, and nothing if the first-period return was low.[4] For the second period, if there is no HWM, then the payoffs are identical to the first period. However, if there is a HWM, it dictates

van Binsbergen (2012) for a related discussion. For evidence of scale diseconomies among hedge funds, see Goetzmann, Ingersoll and Ross (2003) and Naik, Ramadorai and Stromqvist (2007).

[2] Risk aversion would complicate the analysis considerably, given the correlation between realized returns and subsequent expected returns; see Brennan (1998) and Xia (2001).

[3] Assuming that the good type's performance without effort is ex-ante identical to that of the bad type is innocuous and avoids introduction of a third return distribution.

[4] The incentive contract is the standard base fee / incentive fee with HWM hedge fund contract, and for simplicity we assume a zero hurdle rate, i.e. the incentive fee applies to any net profits on an investment, not the net profits in excess of a hurdle rate such as LIBOR.

that any first-period losses suffered by an account must be made up in the second before the incentive fee can be charged to that account. This does not apply to any new date 1 investment in the fund, which will have a fresh HWM and thus pay the incentive fee on positive net profits.. If investment is added to the fund at date 1, the new investment and old investment share in the fund's pre-incentive-fee return in proportion to their invested amounts, and then pay incentive fees out of their take. If investors remove a fraction of their investment before rolling over, the HWM on the remaining investment goes down by the same fraction. All second-period investment pays the base fee $x$ at date 1.

The timing of the game is as follows.

$t = 0$. Nature chooses the manager's type according the prior distribution $\lambda$. The manager privately learns his type. If his type is $G$ then he chooses his effort level; 0 or 1. He announces a contract $(x, y, d) \in (0, 1)^2 \times \{0, 1\}$, where $d = 1$ stands for a HWM and $d = 0$ stands for no HWM. The investors invest. The manager receives his first-period base fee.

$t = 1$. Initial return is realized, $\sigma$ or $-\sigma$. If applicable, the manager receives his first-period incentive fee. The investors make their investment decisions for the second period. The manager receives his second-period base fee.

$t = 2$. The second return is realized, $\sigma$ or $-\sigma$. If applicable, the manager receives his second-period incentive fee.

We use perfect Bayesian equilibrium (PBE) as our equilibrium concept with a focus on efficient equilibria, i.e., equilibria where type $G$ makes the effort and every value-increasing manager manages funds in both periods.

## B. Equilibrium

We start by characterizing the equilibrium set in which efficient effort and management occur. By assumption, effort by type $G$ is efficient, and effort by type $B$ has no effect. Whether it is efficient for $B$ to manage at all depends on $\phi$: if $\phi \geq \frac{1}{2}$, then it is first best efficient to have both types manage in both periods, but if $\phi < \frac{1}{2}$, then it is only efficient if only type $G$ manages. However, the following result shows that it is not possible to achieve first best efficiency when $\phi < \frac{1}{2}$. In particular, type $G$ cannot separate himself from type $B$ while making the costly effort and managing.

**Proposition 1** *Every equilibrium that involves type $G$ making the costly effort and investing is a pooling equilibrium.*

Note that the above result applies for any contract, regardless of whether a HWM is employed. That is, the HWM has no effect on the ability of type $G$ to separate from type $B$. With the manager extracting all surplus, it is just too lucrative for type $B$ not to imitate. One implication is that there would be no investment if the total expected value created by both types, i.e. $\lambda\left(2\left(2\theta-1\right)\sigma-c\right)+(1-\lambda)2\left(2\phi-1\right)\sigma$ is negative. So to keep the analysis interesting we assume that this quantity is positive for the rest of the analysis.

To economize on the exposition, it is useful to introduce some more notation for the further analysis of the pooling equilibirum. After the first period's return, investors update their prior probability $\lambda$ that the manager is type $G$. Let $\lambda_H$ and $\lambda_L$ be the updated probabilities after a high and low return, respectively. From Bayes' rule we have $\lambda_H = \frac{\lambda\theta}{\lambda\theta+(1-\lambda)\phi}$ and $\lambda_L = \frac{\lambda(1-\theta)}{\lambda(1-\theta)+(1-\lambda)(1-\phi)}$. Similarly, investors update their prior probability that the manager will attain the high return. Let $\pi_0$ stand for the manager's unconditional probability of a high return, i.e. $\pi_0 = \lambda\theta+(1-\lambda)\phi$; the probabilities after a high and low return are, respectively, $\pi_H = \lambda_H\theta+(1-\lambda_H)\phi$ and $\pi_L = \lambda_L\theta+(1-\lambda_L)\phi$ . It is also convenient to refer to the investment in the fund at the end of the first period as the fund's *incumbent* investment, and any money added to the fund for the second period as the fund's *new* investment. Next, we first characterize the solution without the HWM, and then with the HWM. This allows us to analyze the impact of HWM.

## C. The No-HWM case

Without a HWM, incumbent and new investment invest on the same terms for the second period, so the Berk and Green (2004) logic applies. It is immediate that their second-period expected returns must be the same, and it also follows that this expectation must be zero. To see that it is zero, note that no investor has anything to gain from accepting a negative expectation, since the second period is the last, and also that if the expectation were positive, new investment would be profitable until it pushed the expectation to zero. This logic applies whether the first–period return is high or low.

From the zero second-period expectation it follows that the first-period expected return is also zero. That is, there is no reason to accept a negative expectation, since the second-period expectation for incumbent investment is certain to be zero, and if the expectaton were positive, investment would enter until the expecation hit zero. Thus, without a HWM, expected returns are zero in all states.

Since expected returns are zero, the fund's assets at the start of each period must be at the level that delivers this expectation. In other words, investment must be at the point where, from the investors' perspective, expected compensation to the manager equals the manager's expected value-added. In the notation developed above, the expected compensation, the base fee plus the expected incentive fee, given investment $I$, is $xI + y\pi_0(\sigma - xI)$ (note that the incentive fee applies to profits net of the base fee), and the manager's expected value-added is $\pi_0\sigma + (1-\pi_0)(-\sigma) = \sigma(2\pi_0 - 1)$. Thus, investment at date 0 is

$$I_0 = \frac{\sigma\left[\pi_0\left(2-y\right)-1\right]}{x(1-y\pi_0)}.$$

Analogously, assets at date 1 after a high first-period return and after a low first-period return are, respectively,

$$\begin{aligned} I_1^H &= \frac{\sigma\left[\pi_H\left(2-y\right)-1\right]}{x(1-y\pi_H)} \\ I_1^L &= \frac{\sigma\left[\pi_L\left(2-y\right)-1\right]}{x(1-y\pi_L)} \end{aligned}$$

since the same break-even logic applies. We assume throughout the paper that we are in the parameter region where these quantities are always positive, which, given that $\pi_L$ is the worst case, amounts to

$$\pi_L > \frac{1}{2-y}.$$

Net fund flows between the two dates are simply the changes dictated by these quantities, and they are in the same or different direction as the first-period return, depending on the magnitude of updating, i.e. the distance between $\pi_0$ and $\pi_H$ or $\pi_L$. If the difference is sufficiently big that the change in break-even fund size is greater than the change in fund size due to returns and fees, then inflows follow high returns and outflows follow low returns. If the difference is sufficiently small, then outflows follow high and inflows follow low. Since a bigger difference means that performance has a bigger effect on beliefs about the manager, it implies lower precision of prior beliefs. So if priors are precise enough then money flows in after low returns; otherwise it flows out.

## D. The HWM Case

It proves to be convenient to divide the HWM case into parameter regions, depending on the direction of net flows after a low first-period return. We define three parameter regions: the *outflow* region, where in equilibrium money flows out after a low return, the *inflow* region, where it flows in after a low return, and the *no-flow* region, where no money flows in or out after a low return. This is the important division because it is after a low return that the HWM matters, and the effects of the HWM are qualitatively different across these regions. The behavior of new flows after a *high* return is not directly interesting, since in that case incumbent investment is at its high water mark, just as new investment would be, so the fund is in exactly the same situation as it is without the HWM, with a zero expected return and assets of $I_1^H$. We first offer some preliminary observations about the qualitative differences between the equlibria in the three regions, and then we proceed to a formal solution of the model.

### D.1. Preliminary Observations

**The Inflow Region** If investors' confidence in the manager is sufficiently high even after a low return, then inflows will partially reverse the reduction in fund size. This new investment invests on the same terms that it would invest on in the no-HWM case, because the HWM can matter only in a subsequent period, but there is no subsequent period. Consequently, it is profitable for new investment to enter after the first period if incumbent investment is less than $I_1^L$, in which case it enters until the fund size reaches $I_1^L$. That is, the expected incentive fee of incumbent investment is irrelevant to the profitability of new investment, so it enters to the same point it would have without the HWM, which by construction is $I_1^L$. Thus, in the inflow region after a low return, the fund's assets are $I_1^L$, new investment enters and earns a zero expected return, and incumbent investment, due to its better terms, stays put and earns a positive expected return.

Because the expected second-period return on incumbent investment is positive, the expected first-period return has to be negative in equilibrium, driven by investment in excess of $I_0$. Effectively, the initial investment delivers both the first-period return *and* an option to roll over as an incumbent after a low return for the positive second-period expectation, so initial investment drives down the expected return until the two expectations offset. So in the inflow region, initial expected returns are negative, and initial investment is greater than it is without the HWM.

**The Outflow Region** If the expected return on incumbent investment is negative after a low first-period return, then it is even more negative for new investment, so new money will not enter, and some incumbent money must flow out for the remaining investment to break even. So money flows out, and the expected return is zero. Thus, the expected second-period return is zero no matter what, which implies that the first-period expected return is zero as well, since there is no reason for investors to accept anything worse.

In the outflow region, initial investment in the fund is the same as without the HWM, but investment after a low return is greater. The initial investment has to be $I_0$, as it is without the HWM, since that is the fund size that delivers a zero first-period expectation. And after a low return, the equilibrium investment must be greater than $I_1^L$, the investment that obtains without a HWM, because $I_1^L$ is the fund size at which new investment just breaks even, whereas the expected return of new investment in this region is negative, as it is worse than the expectation of incumbent investment, which is 0.

**The No-Flow Region** The remaining possibility is that the expected return on incumbent investment is positive, but the expected return on new investment is negative. In this case it is unprofitable to either add or subtract investment, so no money flows in or out. Consequently, the expected return on incumbent investment is positive in the second period after a bad return, and the expected first-period return is negative. From the analysis of the other regions, it follows that the fund is larger than in the no-HWM case both initially and after a low return.

### D.2. Solving the Model

We solve the model by identifying the boundaries between the parameter regions, and for each region, calculating expected returns and fund sizes in both periods. The boundaries between the regions are discovered by comparing three quantities: the break-even initial fund size with no HWM, defined as $I_0$ above; the break-even initial fund size *with* a HWM but *without* inflows or withdrawals after a bad initial return, which we denote $I_0^S$; and the initial fund size such that the investor breaks even after a bad return, without inflows or withdrawals at date 1, assuming no HWM. Since $I_1^L$ is by construction the break-even fund size after a bad return, assuming no HWM, it follows that this final quantity is $\frac{I_1^L+\sigma}{1-x}$. The functional form of $I_0^S$ is the solution to the equation

$$\begin{aligned}&-xI_0^S + \pi_0(\sigma - y(\sigma - xI_0^S)) + 0\pi_0 + (1-\pi_0)(-\sigma)\\&+(1-\pi_0)[-x(I_0^S(1-x)-\sigma) + \pi_L(\sigma) + (1-\pi_L)(-\sigma)] \quad = 0\end{aligned}$$

which rearranges to

$$I_0^S = \frac{\sigma[\pi_0(2-y) + (1-\pi_0)(2\pi_L + x - 1) - 1]}{x[(1-\pi_0 y) + (1-\pi_0)(1-x)]}.$$

The following propositions identify the parameter regions and the expected returns and investment levels in those regions, which we distinguish from the no-HWM case by adding a hat. That is, the HWM analogs to $I_0$, $I_1^H$ and $I_1^L$ are $\widehat{I}_0$, $\widehat{I}_1^H$ and $\widehat{I}_1^L$, respectively. We already know that $\widehat{I}_1^H = I_1^H$ and that expected returns after a high return are zero in all regions so the propositions focus on the other quantities.

**Proposition 2** *We are in the Outflow region if and only if $I_0^S \leq I_0$. In this region, expected returns are* 0 *in both periods and HWM has no impact on the initial fund size, $\widehat{I}_0 = I_0$, but leads to higher fund size in the second period following a low return, $\widehat{I}_1^L > I_1^L$.*

Expected first-period profits are not tenable in equilibrium, so $\widehat{I}_0$ must be at least $I_0$. So if $I_0^S \leq I_0$, then incumbent investors must face expected losses if they leave all their money in after a bad return, because they will have invested at least $I_0^S$ in the first period, and investing $I_0^S$ implies first-period expected profits and thus expected losses from leaving money put after a bad return. Expected second-period losses are not tenable in equilibrium, so they will take money out until they break even. Since they break even in the second period, they must break even in the first as well.

**Proposition 3** *We are in the inflow region if and only if $I_0 < I_0^S < \frac{I_1^L+\sigma}{1-x}$. In this region, the expected return in the first period is negative whereas the expected second-period return is positive. HWM has no impact on the second-period fund size, $\widehat{I}_1^L = I_1^L$, but leads to higher fund size initially $\widehat{I}_0 > I_0$.*

With $I_0^S > I_0$, we must be in either the Inflow or No-Flow region. To see that the additional inequality $I_0^S < \frac{I_1^L+\sigma}{1-x}$ puts us in the Inflow region, note that by definition this implies that if the initial investment is $I_0^S$, then the amount left in the fund after a bad return will be below $I_1^L$, which implies positive expected returns on inflows until the fund size reaches $I_1^L$, while by the same token, $I_0^S \geq \frac{I_1^L+\sigma}{1-x}$ would imply a non-positive expected return on inflows. So in the latter case, $\widehat{I}_0 = I_0^S$ and there are no flows after a bad return, and in the former case, there are inflows after a bad return, and $I_0 < \widehat{I}_0 < I_0^S$ (reducing first-period expected losses, though not to zero, to offset the dilution by inflows of expected second-period profits). So we can also characterize the No Flow region:

**Proposition 4** *We are in the no-flow region if and only if $I_0^S > \max\{I_0, \frac{I_1^L+\sigma}{1-x}\}$. In this region, the expected return in the first period is negative whereas the expected second-period return is positive. HWM leads to higher fund size both initially and in the second period following a low return, $\widehat{I}_0 = I_0^S > I_0$ and $\widehat{I}_1^L = I_0^S(1-x) - \sigma > I_1^L$.*

### D.3. Comparative Statics

In the no-HWM case, the relation of flows to performance turns from positive to negative as precision increases. In the HWM case this is still broadly true, That is (by inspection of its functional form), $I_0^S$ increases with $\pi_L$, holding $\pi_0$ constant, and therefore increases with precision, and the propositions show that $I_0^S \leq I_0$ is necessary for money to flow out after a bad return, and $I_0^S > I_0$ is necessary for money to flow in. But the HWM also introduces an intermediate region with no flows, and it also alters expected returns. We can summarize the effect of the HWM within each parameter region on expected returns and fund size, and by implication on net flows, in a table. Here each entry is the net effect of the HWM in the sense that the value under a contract without a HWM is subtracted from the value under an otherwise identical contract *with* a HWM. We use + and - to represent positive and negative net effects, respectively, and 0 to represent no net effect.

The effect of the HWM varies across the regions. Since the regions depend on precision, it follows that the effect of the HWM on quantities of interest depends on precision. To identify this dependence, we present a series of figures where we hold all parameters constant except $\lambda$ and $\theta$, which we vary to hold $\pi_0$ constant while increasing $\pi_0 - \pi_L$, i.e., decreasing precision. So in each figure, the investors' initial prior on a high return stays fixed while, from left to right, the sensitivity of this prior to a low return increases.

Figure 1 shows the effect of decreasing precision on the fund's initial assets. As precision falls, the economy moves from the inflow to no-flow to outflow regions, and assets consequently fall, leveling off at the size of a fund with no HWM. So the more precisely investors know a new manager's ability, the more they invest. They consequently accept lower initial expected returns and enjoy higher expected returns after poor performance, as Figure 2 illustrates. Thus, initial expected returns decrease and fund size increases, as prior precision increases. Similarly, after a low return, both positive and zero flows predict higher performance than do outflows. Also, expected future returns are higher after low than after high returns, but only when prior precision is high, which is to say, only when low returns beget inflows or no flows, rather than outflows. In taking these results to the data, it would be worth bearing in mind that it is the incumbent money, rather than incoming money, with the positive expectations. Empirical work on hedge-fund returns is typically silent about which investment's return is being addressed, but this distinction is crucial.

In Aragon and Qian (2010), the HWM is shown to encourage investors to stay when they would have left. Table 1 shows that the HWM performs this duty in the no-flow and outflow, but not inflow regions. This is illustrated in Figure 3, which is analogous to Figures 1 and 2 except that it addresses fund size after a low return. Without a HWM, fund size declines linearly. With a HWM, fund size departs from this relation when the model enters the no-flow region, and the gap remains through the outflow region, though it starts to contract when precision gets very low. This contraction at the extreme arises from investors withdrawing so much that they reintroduce the possibility of paying an incentive fee due to the shrinking of the HWM. That is, if investors leave all their money in the fund, then the fund must earn back its entire first-period loss before an incentive fee is paid.. But if investors withdraw, say, 2/3 of their money, then the fund must earn back only 1/3 of its first-period loss before an incentive fee is paid. Thus, drawdowns can potentially increase expected incentive fees. Results for extreme drawdowns are more sensitive to the effect of reduced capital on the fund's ability to add value, which this model abstracts from.

Net flows are usually discussed in percentage terms, i.e. divided by the fund's assets just before the flows. Accordingly, Figure 4 divides the dollar flows in Figure 3 by the relevant end-of-first-period assets. The notable result here is that the effect of the HWM on percentage flows is actually negative in the inflow region, due to the inflation of initial assets caused by the HWM, and that the effect is non-monotonic, increasing through-the no-flow region, and then decreasing again in the extreme outflow region discussed above.

The HWM also has an effect on flows after high returns, due to the effect on initial investment in Figure 1. This effect, illustrated in Figure 5, attenuates inflows of the funds where precision was higher in the first place.

# II. Efficiency

We have already seen that the HWM does not address the adverse selection problem, because it does not discourage low-ability types from imitating high-ability types. The remaining question is whether the HWM addresses moral hazard, by increasing what high-ability types get from making the value-increasing effort. The HWM has no effect after a high return, so its effect must come from the initial period, or the period subsequent to a low return. With respect to first-period expected fees, the HWM has no direct effect but it can have an indirect effect, because in the inflow and no-flow regions it increases the initial fund size. However, the resulting effect on the incentive to make an effort can only be negative, because the increased base fee comes out of the profits that the incentive fee applies to and thus decreases the reward in first-period fees from making an effort. With respect to second-period fees after a low return, the nature of the HWM's effect on incentives depends on the parameter region. In the outflow region, the HWM does not change expected fees to the manager, from the perspective of investors who break even with or without it; it merely shifts fees from incentive to base. Intuitively, this shift is bad for the reward for effort. In the other regions, the HWM reduces expected fees to the manager after a low return, as evidenced by incumbent investors' positive expected returns. This reduction is good for the reward for effort, since it reduces the payoff from getting a low return in the first place, but there is a countervailing incentive effect from the fact that this reduction comes from reducing incentive fees.

We address the efficiency implications of the HWM by asking whether, for a given base and incentive fee, adding the HWM increases or decreases the maximum effort cost that a type $G$ willingly pays in equilibrium. We proceed by first establishing separately for the no-HWM and HWM cases that, for a given $x$ and $y$, there is a cutoff effort cost such that an equilibrium where type $G$ makes an effort exists if and only if $c$ is below this cutoff. We then address the efficiency effect of the HWM in each region by determining whether this cutoff is higher or lower when the HWM is added.

The constraint $\pi_L > \frac{1}{2-y}$,which ensures positive investment even after a low return, rearranges to an upper bound on the incentive fee: $\overline{y} = 2 - \frac{1}{\pi_L}$. Thus, the analysis is limited to incentive fees below this ceiling. We begin this analysis by characterizing the set of efficient equilibria without the HWM.

**Proposition 5** *For any given $x \in (0,1)$ and $y \in (0,\overline{y})$, there exists a $\overline{c}(x,y,d=0)$ such that there exists an efficient pooling equilibrium if and only if $c \leq \overline{c}$.*

When the effort cost is sufficiently low, i.e. $c \leq \overline{c}$, type $G$ pays the cost and both types offer a contract without a HWM. The potential improvement by the HWM is that it may generate an efficient outcome even when $c > \overline{c}$. In order to see whether this is possible, we characterize the set of efficient equilibria with a HWM.

**Proposition 6** *For any given $x \in (0,1)$ and $y \in (0,\overline{y})$, there exists a $\widehat{c}(x,y,d=1)$ such that there exists an efficient pooling equilibrium if and only if $c \leq \widehat{c}$.*

Next, we analyze the impact of adding the HWM by determining whether $\widehat{c}$ is higher than $\overline{c}$, in which case the HWM improves efficiency by increasing the highest effort cost that a type $G$ willingly pays, or if instead it is lower, in which case it reduces efficiency.

It is worth noting that, as is common in complete-information games, the model admits multiple equilibria, and we select and focus on the most efficient equilibria with and without the HWM. Alternatively, we could use standard refinements such as Perfect Sequential Equilibrium, by which a contract with a HWM is a more robust equilibrium outcome if and only if it is more efficient than the one without the HWM. We analyze the regions separately, starting with the Inflow case.

## A. Inflow Case

In the Inflow case, whether or not the HWM increases efficiency hinges on a simple inequality among the parameters:

**Proposition 7** *In the Inflow Region, for any given* $x \in (0,1)$ *and* $y \in (0,\overline{y})$ *a high water mark contract leads to a higher incentive to make costly effort, i.e.,* $\widehat{c}(x,y,d=1) > \overline{c}(x,y,d=0)$*, if and only if* $\theta + \phi > 1 + y\theta\phi$*.*

We can gain some of the intuition behind this result by considering the three ways the HWM affects efficiency in this region. Through the effect of effort on the chance of a good first-period return, the HWM brings both a cost and a benefit. The cost arises from the extra base fee from the extra initial investment, because the extra base fee reduces the incentive fee by reducing the net profits it applies to. The benefit arises from the positive expected return of incumbent investment after a bad return, which comes out of the manager's expected compensation after a bad return, thereby increasing the manager's punishment for earning the bad return. Through the effect of effort on the chance of a good return after a bad return, the HWM brings only a cost, because it reduces the manager's incentive pay after a bad return. Looking at just this last effect, from the perspective of a type $G$ choosing at date 0 whether to make the effort, this effort raises the probability of a bad return followed by a good return by $(1-\theta)\theta - (1-\phi)\phi$, which is negative if and only if $\theta + \phi > 1$. That is, making an effort decreases the probability of suffering this cost if and only if $\theta + \phi > 1$. That the first cost increases with $y$ helps explain why it raises the lower bound on $\theta + \phi$. An exact derivation of the functional form is in the proof.

One implication of this result is that the efficiency of the HWM depends on the quality $\phi$ of the bad type. To see this, note first that the inequality in the proposition rearranges to $\theta > (1-\phi)/(1-y\phi)$, which if $\phi = 1/2$, i.e. the bad type's value added is simply 0, reduces to $\theta > 1/(2-y)$. This has to hold, because we are already assuming $\pi_L > \frac{1}{2-y}$, and because $\theta > \pi_L$. So in a high-precision economy where money enters after a bad return, if a bad manager adds zero value then the HWM always improves efficiency. Since $(1-\phi)/(1-y\phi) < 1/(2-y)$ if and only if $\phi > 1/2$, the HWM also always adds value if the bad manager adds positive value, but by the same token if $\phi < 1/2$, i.e. a bad manager subtracts value, there is a parameter range, $(1-\phi)/(1-y\phi) > \theta > \pi_L > 1/(2-y)$, where the HWM reduces efficiency.

## B. Outflow Case

In a low-precision economy where money flows out ofter a bad return, the HWM always reduces efficiency:

**Proposition 8** *In the Outflow Region, for any given $x \in (0,1)$ and $y \in (0, \overline{y})$, a high water mark contract leads to lower incentive to make costly effort, i.e., $\widehat{c}(x, y, d=1) < \overline{c}(x, y, d=0)$.*

In the Outflow Region, the HWM does not affect first-period fund size, and so does not affect first-period compensation. After a low initial return, the HWM makes the fund larger, so it increases the base fee. Therefore, a good type making an effort calculates an expected gain of $(1-\theta)$ times this increase, and a bad type or shirking good type calculates an expected gain of $(1-\phi)$ times this increase. However, the base-fee increase accompanies an incentive-fee decrease. The good type making an effort puts probability $(1-\theta)\theta$ on this decrease, whereas the bad type and shirking good type put probability $(1-\phi)\phi$. From the investors' point of view, the expected increase must offset the expected decrease, given that they break even overall, but a manager's expectation is different because he knows his type. Since $\frac{(1-\theta)\theta}{(1-\theta)} > \frac{(1-\phi)\phi}{(1-\phi)}$, it follows that the net effect is negative for good types and positive for bad types and shirking good types. So in the Outflow region the HWM makes it less attractive to incur costs to become a better money manager.

## C. No-flow Case

In the No-Flow Region, between the other two, the result is similar to the Inflow result, but the condition identifying the region where the HWM increases efficiency is not as tidy:

**Proposition 9** *In the No-Flow Region, for any given $x \in (0,1)$ and $y \in (0,\overline{y})$ a high water mark contract leads to a higher incentive to make costly effort, i.e., $\widehat{c}(x,y,d=1) > \overline{c}(x,y,d=0)$, if and only if $(\theta+\phi-1-\pi_L)\, y\frac{(1-\pi_L)}{(1-y\pi_L)} > \frac{(1-y)(\pi_L(1-y\pi_0)-\pi_0(1-x)(1-y))}{(1-y\pi_0)((1-\pi_0 y)+(1-x)(1-\pi_0))}$.*

This expression doesn't allow much intuition. There is some resemblance to the Inflow result in that $\theta+\phi-1$ must be sufficiently large, but beyond that the model's dynamics are hard to tease out. Accordingly, the next section numerically explores when the HWM provides efficiency benefits and outlines the empirical predictions one can take to the data.

# III. Discussion

## A. Effectiveness of the HWM

To explore the implications of Propositions 9 and **7** for efficiency, we numerically evaluate the parameter regions where the HWM alleviates moral hazard. For this purpose we hold the base and incentive fees to the standard 2% and 20%, the bad manager's value added to zero (i.e.$\phi = 0.5$), and the return $\sigma$ to 100 (just to scale the problem), while varying the incidence $\lambda$ and ability $\theta$ of the good manager. Figure 6 plots the three flow regions, and Figure 7 shows where the HWM does or does not add value.

Figure 6 demonstrates the high $\lambda$ required for the Inflow region. In this parameter space, $\lambda$ has to be almost 99%. The figure also demonstrates that the range of $\theta$ supporting this region is narrow and bounded away from one. It stops short of one, even though very high $\theta$ means the good type is very good, because a tiny value for $1 - \theta$ makes a bad return very bad news, thus encouraging outflows. Consequently, $\theta$ has a non-monotonic effect, with both low and very high values ruling out inflows after bad returns.

Unlike the boundary between No Flow and Inflow, the boundary between No Flow and Outflow is monotonic in $\theta$. In this moderate range of $\theta$, its effect on the quality of the unconditional pool of managers outweighs the conditioning effect of a bad return. The effect of $\lambda$ is more straight-forward. Increasing $\lambda$ moves the model from Outflow to No Flow and then from No Flow to Inflow.

Regarding moral hazard, we already know that the HWM makes it worse in the Outflow region, and better in the Inflow region (provided that, as the figures assume, the manager's value-added is zero) What Figure 7 shows us is how the No Flow region bridges the two. In this region, high $\lambda$ improves efficiency simply by moving the model closer to the Inflow region. The impact of $\theta$ is, however, more subtle. To understand its effect on efficiency, first, note that the HWM increases fund size both in the first period and in the second period after a bad return. The larger initial size means a larger initial base fee, which in turn means a lower initial incentive fee for a high initial return. So a higher $\theta$ means a higher expected loss of this first-period incentive fee, and note that this expected loss is linear in $\theta$. If the first-period return is bad, in the Inflow region the HWM both raises the second-period base fee and eliminates the second-period incentive fee, where the base-fee gain and incentive-fee loss are realized with probabilities $(1-\theta)$ and $(1-\theta)\theta$, respectively, for the good type making an effort, and probabilities $(1-\phi)$ and $(1-\phi)\phi$, respectively, for the bad type and shirking good type. Thus, the HWM's effects on first- and second-period base fees are both disincentives for effort, and both linear in $\theta$, whereas the effect the effect on the second-period incentive fee is quadratic in $\theta$. And with $\phi = 0.5$, implying $(1-\phi)\phi = 0.25$, and also $\theta > \phi$, it follows that $(1-\theta)\theta < 0.25$, and also that increasing $\theta$ further lowers $(1-\theta)\theta$, and at an increasing rate. Therefore, higher $\theta$ means higher incentive to make the effort, as the benefit from reducing the probability of the reduced incentive fee grows relative to the base-fee effects. Figure 7 illustrates the non-linearities in the relation between efficiency and $\theta$ and most importantly shows that only in a small region, all lying above $\lambda > 0.91$, does the the HWM encourage, rather than discourage, costly effort. These results cast doubt on the effectivenss of the high-water mark and highlights the limits of this type of contract.

## B. Empirical Predictions

The comparative statics illustrated in Figures 1 through 5 are empirically testable, to the extent one can proxy for the quantity on the x-axis, the sensitivity of the manager's alpha to bad performance. As this sensitivity is essentially the imprecision of ex ante beliefs about the alpha, and since imprecision tends to decline with the quantity of data, an intuitive proxy for this quantity is the length, or more precisely, the shortness, of the manager's track record. So as the track record shortens, the model predicts the marginal effects of the HWM in the figures:

- Initial assets shrink, then flatten out (Figure 1)
- Initial returns grow, then flatten out (Figure 2)
- Returns on incumbent investment after poor performance decline, then flatten out (Figure 2)
- Assets after poor performance grow, then flatten out (Figure 3)
- Net flows after good performance grow, then flatten out (Figure 4)
- Net flows after poor performance grow, then flatten out (Figure 5)

Another perspective on the same figures is that they relate not only imprecision but also flows after bad performance to the quantities on the y-axes. So the prediction from Figure 2 is that, after bad performance, the marginal effect of the HWM is higher returns for incumbent investment in the case of inflows or zero flows. In the case of outflows, the HWM has no marginal effect on returns. Similarly, Figures 3 and 5 show that the positive effect of the HWM on assets after a bad return is greater in the case of outflows than of inflows or zero flows, and Figure 4 shows that, after good performance, the HWM reduces net flows, except in the case of outflows.

Regarding efficiency, the main prediction is the negative result about adverse selection, i.e. that the presence of a HWM is unrelated to the presence of skill, so the HWM has no marginal effect on any measure of the manager's value-added. The moral-hazard result, i.e. that the HWM adds value only when it is highly probable that the manager is skilled, can be taken as a prediction that the HWM will be used more often when imprecision is very low.

## C. Empirical Studies

The empirical literature has not tested the predictions about precision, but some other findings are relevant. Two recent empirical studies of hedge-fund fees, Agarwal and Ray (2012) and Deuskar, Wang, Wu, and Nguyen (2012), find about twice as many funds with HWMs than without, and in the cross-section, neither study finds any relation in the cross section between the presence of HWMs and performance. This is consistent with the HWM not correlating with skill, as the model predicts, but since the model also predicts that fund size adapts to equalize performance, it would also be interesting to know if the HWM is unrelated to fund size.

More to the point of our analysis is the finding by Agarwal and Ray (2012) that returns are negative (significantly or not, depending on the test design) in the first six months after the introduction of a HWM, a period that corresponds roughly to the first period of our model. Combined with the negative result for the unconditional effect of the HWM on returns, this bears out the prediction that investors buy the chance at a positive future expected return with a negative initial expected return. They also find that returns before the HWM introduction are abnormally high, which is at least in the direction of the prediction that the HWM adds value only for managers whose skill is close to certain.

In Goetzmann, Ingersoll and Ross(2003), one of the main empirical findings is that, subsequent to good performance, small funds grow and large funds shrink. This does not address the predictions of our model but it is worth noting, as it is consistent with the large funds, but not small funds, being in the inflow region, which is where the efficiency of the HWM is relatively high. The assumption of decreasing returns to scale and competitive capital markets in hedge funds was recently tested by Naik, Ramadorai and Stromqvist (2007) and Fung, Hsieh, Naik and Ramadorai (2008), who conclude that the effect is strong in recent years, and weaker in earlier years.

# IV. Concluding Remarks

The HWM governs the fees of many funds, which hold a large portion of managed wealth, so it is important to understand both why it exists and what it does. The literature to date has explored the HWM along several dimensions, but has simplified the analysis to that of one investment that might eventually leave. In this paper we expand the analysis to that of a fund that not only lets money out but takes it in as well. Because the HWM applies at the investment level, this can differentiate the expected fees, and thus expected returns, of the investors within the fund, and we find this differentiation to be key to the role of the HWM. Most notably, to those entering the fund of a well-known manager, it offers the prospect of rents after bad performance, and it does this by shielding incumbent investors from the willingness of new investment to compete its own rents to zero. To those entering the fund of a rookie manager, the HWM means little, mainly because the much-worse prospects after bad performance mean there are no rents to shield. Thus, the effects of the HWM include distortions of the time series and cross section of funds' returns, and also of their assets under management, where the crucial distinction between funds is not the alphas investors initially ascribe to the managers, but the precisions of these alphas.

We analyze performance of HWM in mitigating the perverse effects of adverse selection and moral hazard. We address adverse selection by allowing managers to be skilled or unskilled, where the skilled managers can try to reveal themselves through the contracts they offer. The HWM, we find, does not help solve this problem, because the unskilled is always better off choosing to offer whatever contract the skilled offer, whether it includes a HWM or not. We also address moral hazard by requiring a privately costly effort to turn skill into value, and while we find that the HWM can reward effort, in the sense of increasing the maximum cost managers will pay, we also find that the parameter region where this increase occurs is unlikely to obtain in practice. For the HWM to add value by alleviating moral hazard, investors would have to put a very high prior probability on the manager being skilled.

If the HWM is so little help against adverse selection and moral hazard, then what problem, if any, does it address? One possibility raised by our analysis is that it accelerates the compensation of experienced managers, which might be Pareto improving if the manager is less patient than the investors. That is, with the HWM, the manager effectively sells an above-market return in certain future states, and enjoys the sale proceeds immediately through the increased base fee it brings in. Since experienced managers would tend to be older, and since many investors are patient capital pools such as endowments, this acceleration could be a welfare-increasing trade. Whether this or some other rationale explains the popularity of the HWM is an interesting area for empirical research.

# Appendix

**Proof of Proposition 1:** First, suppose that such a separating equilibrium exists. Then there must exist a pair of contracts $(x, y, d) \neq (x, y, d)'$ such that type $G$ offers $(x, y, d)$ and type $B$ offers $(x, y, d)'$. Recall that a manager must extract the entire expected surplus at date 0 in equilibrium. So in a separating equilibrium the expected payoffs are $2\sigma(2\theta - 1) - c$ and $\max\{2\sigma(2\phi - 1), 0\}$ for types $G$ and $B$, respectively. Next, we check for deviations by type $B$. If type $B$ deviates and offers $(x, y, d)$ then the investors believe that the manager's type is $G$ with probability 1. If instead the investors had believed that the manager's type had been $B$ with probability 1, then the expected payoff of a type $B$ manager would have been $2\sigma(2\phi - 1)$ under the $(x, y, d)$ contract. Note that for a given contract the initial fund size is strictly greater and the second period's fund size is weakly greater when investors assign higher probabilities to type $G$, so type $B$'s expected payoff is strictly greater than $2\sigma(2\phi - 1)$ when he deviates and offers $(x, y, d)$. Therefore, there does not exist an efficient separating equilibrium. A similar argument rules out a partially separating equilibrium as well.

**Proof of Proposition 2:** We have already shown that $\widehat{I}_0 = I_0$ in the outflow region, and since in this region investors break even in the first period and must withdraw investment to break even after a low return, it follows that their initial investment is too high to break even overall without withdrawing. Thus in the outflow region, it must be true that $I_0^S < I_0$. For the other direction, note that if $I_0^S < I_0$ then investors will not invest more than $I_0$ in the first period, since that way their expected return is negative in the first period and zero in the second, since they will have to withdraw money after a low return just to break even. And they will not invest less than $I_0$ because then it is profitable to invest for just the first period. So they invest $I_0$ and then withdraw some investment after a low return. $\Diamond$

**Proof of Proposition 3:** If $\widehat{I}_0 = I_0^S > I_0$ then second-period expected returns are positive, and if $I_0^S < \frac{I_1^L+\sigma}{1-x}$ then new investment will enter, bringing down the profitability of incumbent investment, thereby pushing the overall expected return as of date 0 below zero. To bring the expected return back to zero, investors must invest less, but investment will not fall all the way to $I_0$, because in that case they would be breaking even in the first period and making expected profits in the second. So there is an $I_0^F$ between $I_0^S$ and $I_0$ at which they break even overall. To identify $I_0^F$, we use the fact that rolled-over investment will share the fund's pre-incentive-fee return with new investment in proportion to their invested amounts, and we use our result from above that the total of rolled-over and additional investment is $I_1^L$. Thus, $I_0^F$ solves

$$-xI_0^F + \pi_0(\sigma - y(\sigma - xI_0^F)) + (1-\pi_0)(-\sigma) + (1-\pi_0)[-x(I_0^F(1-x) - \sigma)$$

$$+\frac{I_0^F(1-x) - \sigma}{I_1^L}(\pi_L(\sigma) + (1-\pi_L)(-\sigma))] = 0,$$

which is strictly greater than $I_0$. To calculate the expected second-period return after a low return, plug $I_0^F$ into the expected second-period return, i.e. $-x[(1-x)I_0^F - \sigma] + (2\pi_L - 1)\sigma$, which gives

$$\frac{2\sigma\pi_L y(1-\pi_L)(\pi_0[2(1+xy-x)-y]-1)}{(1-\pi_0 y)(\pi_L(2-y)-1)-(1-x)(1-\pi_0)2\pi_L y(1-\pi_L)} > 0.$$

For the overall expected return to be zero, the expected first-period return must be $-(1-\pi_0)$ times this number. $\Diamond$

**Proof of Proposition 4:** The no-flow region requires positive expected second-period expected returns from rolling over, which is true if and only if $I_0^S > I_0$. It also requires negative expected second-period returns on additional investment. This is true if and only if $I_0^S(1-x) - \sigma > I_1^L$, because this implies that the rolled-over amount, $I_0^S(1-x) - \sigma$, is greater than the break-even investment level at the HWM of new investment. Thus we also have $\widehat{I}_1^L = I_0^S(1-x) - \sigma$. To calculate the expected second-period return after a low return, plug $I_0^S$ into the expected second-period return, i.e. $-x[(1-x)I_0^S - \sigma] + (2\pi_L - 1)\sigma$, which gives

$$\frac{2\sigma[\pi_L(1-\pi_0 y) - \pi_0(1-y)(1-x)]}{(1-\pi_0 y) + (1-\pi_0)(1-x)} > 0.$$

For the overall expected return to be zero, the expected first-period return must be $-(1-\pi_0)$ times this number, therefore it is strictly less than zero. $\Diamond$

**Proof of Proposition 5:** First we assume that such an equilibrium exists and use Bayesian updating on the equilibrium path. Next, we write down the incentive compatibility condition for type $G$ making the costly effort.

$$\begin{aligned}
& xI_0 + \theta\left(y\left(\sigma - xI_0\right) + xI_1^H + \theta y\left(\sigma - xI_1^H\right)\right) + (1-\theta)\left(xI_1^L + \theta y\left(\sigma - xI_1^L\right)\right) - c \\
\geq\; & xI_0 + \phi\left(y\left(\sigma - xI_0\right) + xI_1^H + \phi y\left(\sigma - xI_1^H\right)\right) + (1-\phi)\left(xI_1^L + \phi y\left(\sigma - xI_1^L\right)\right)
\end{aligned}$$

Simplifying terms lead to

$$(\theta-\phi)\left(y\left(\sigma-xI_0\right)+y\left(\sigma-xI_1^L\right)+xI_1^H-xI_1^L\right)+\left(\theta^2-\phi^2\right)xy(I_1^L-I_1^H)\geq c$$

Further simplifying we have

$$(\theta-\phi)\left(2y\sigma-xyI_0-xyI_1^L+xI_1^H-xI_1^L\right)+\left(\theta^2-\phi^2\right)xy(I_1^L-I_1^H)\geq c$$

Solving for the cutoff cost, $\overline{c}$, we have

$$\overline{c}=(\theta-\phi)\left(2y\sigma+x\left(I_1^H-I_1^L\right)-xy(I_0+I_1^L)\right)-\left(\theta^2-\phi^2\right)xy(I_1^H-I_1^L).$$

Next, we characterize the equilibrium and check for deviations. In equilibrium, type $G$ makes the costly effort. Both types offer $(x, y, d=0)$ and investors' initial beliefs are equal to the common priors. They invest a total amount of $I_0=\frac{\sigma[\pi_0(2-y)-1]}{x(1-y\pi_0)}$. The manager receives $xI_0$ and $xI_0+y\left(\sigma-xI_0\right)$ after a low and high return respectively. The investors update their beliefs to $\lambda_L$ and $\lambda_H$, and invest a total amount of $I_1^L=\frac{\sigma[\pi_L(2-y)-1]}{x(1-y\pi_L)}$ and $I_1^H=\frac{\sigma[\pi_H(2-y)-1]}{x(1-y\pi_H)}$ following low and high return respectively. The second-period returns are realized and managers receive $xI_1$ and $xI_1+y\left(\sigma-xI_1\right)$ after a low and high return, respectively. For complete characterization of the entire parameter space, we impose the off the equilibrium path beliefs that allow us to support the largest set of equilibria. In particular, we complete the characterization with the following: if the manager offers any other contract, $(x', y', d')$, investors believe that manager's type is $B$.

Given these beliefs, if $\phi\leq\frac{1}{2-y'}$ then deviation leads to zero total fees since the fund size is zero in both periods. So deviation may be profitable only if $\phi>\frac{1}{2-y'}$. Given that there is no updating, if the second period fund size is positive, the first period fund size is also positive. Note that the expected total fee for type $B$ is equal to $2\sigma\left(2\phi-1\right)$. Recall that type $B$ would have made the same amount under the equilibrium contract

$(x, y, d = 0)$ if the investors assign probability 1 to having type $B$. Therefore, the equilibrium payoff has to be higher for type $B$ given that the beliefs put strictly higher probability on type $G$ on the equilibrium path. Consequently type $B$ cannot deviate profitably. Similarly, one can verify that there is no profitable deviation for type $G$ at the offer stage,either. Finally, by the definition of $\bar{c}$ a deviation by type $G$ is not profitable at the effort stage as long as $c \leq \bar{c}$. $\Diamond$

**Proof of Proposition 6:** The proof is omitted to avoid repetition. $\Diamond$

**Proof of Proposition 7:**As discussed in the previous section, the HWM increases the initial fund size in the Inflow Region, i.e., $\widehat{I_0} > I_0$. The expected second-period cost of the HWM (in terms of loss of incentive fee) for the manager, and conversely, the second-period gain for incumbent investors, is $(1-\pi_0)\pi_L r y \left(\sigma - x I_1^L\right) > 0$, where $r = \min\{\frac{\widehat{I_0}(1-x)-\sigma}{I_1^L}, 1\}$ (note that $r$ is the fraction the fund's second period assets that incumbent investment represents). To offset the second-period expected benefit with greater first-period fees, the fund is larger at date 0 with the HWM, i.e. $\widehat{I_0} > I_0$. Let $\Delta I_0$ denote $\widehat{I_0} - I_0$. There are two channels through which compensation is affected. First, higher fund size leads to higher base fees which leads to equally higher payoffs for both types $B$ and $G$ in comparison to the no-HWM case. Second, higher base fees, $x\widehat{I_0}$, means smaller incentive fees, $\pi_0 y \left(\sigma - x\widehat{I_0}\right)$. Therefore, due to the HWM expected compensation in the initial period increases by $x\Delta I_0 - \pi_0 y x \Delta I_0$ which can be rewritten as $x\Delta I_0(1-\pi_0 y)$. This higher payoff for the manager in the first period is not equally shared by both types. Smaller incentive fees affect type $G$ more negatively since $\theta y x \Delta I_0 > \phi y x \Delta I_0$. Therefore, type $B$ expects to benefit more in the initial period due to the HWM.

Both types pay in the second period for their extra compensation in the first: the net expected gain in first period, $x\Delta I_0(1-\pi_0 y)$, has to equal to the net expected loss of incentive fees in the second period, $(1-\pi_0)\pi_L ry\left(\sigma - xI_1^L\right)$. Furthermore, the zero expected compensation change implies that

$$(1-\pi_0)\pi_L ry\left(\sigma - xI_1^L\right) = x\Delta I_0(1-\pi_0 y)$$

which can be rearranged to

$$ry\left(\sigma - xI_1^L\right) = x\Delta I_0 \frac{(1-\pi_0 y)}{(1-\pi_0)\pi_L}.$$

Since the expected effect of the HWM on compensation depends on realized returns, it varies across the types. For type $G$, the expected second-period cost of the HWM is $(1-\theta)\theta ry\left(\sigma - xI_1^L\right) > 0$, and for type $B$ it is $(1-\phi)\phi ry\left(\sigma - xI_1^L\right) > 0$. Note that efficiency requires $\theta > \frac{1}{2}$, so if $\theta + \phi < 1$, we must have $1-\phi > \theta > 1-\theta > \phi$, which implies $(1-\theta)\theta > (1-\phi)\phi$. Therefore, $(1-\theta)\theta ry\left(\sigma - xI_1^L\right) > (1-\phi)\phi ry\left(\sigma - xI_1 I_1^L\right)$. Thus, if $\theta + \phi < 1$, then type $G$ has a higher expected second-period cost due to HWM. Recall that type $G$ also has a lower expected benefit in the first period. Since there is zero net expected gain across periods and types, the HWM must make type $G$ worse off, and type $B$ better off, when $\theta + \phi < 1$. So the expected compensation difference between the types is smaller, and consequently there is less incentive for type $G$ to make the costly effort.

On the other hand, if $\theta + \phi > 1$, then $\theta > \max\{1-\phi, \phi\}$ and $\min\{1-\phi, \phi\} > 1-\theta$, so $(1-\theta)\theta < (1-\phi)\phi$. Therefore, $(1-\theta)\theta ry\left(\sigma - xI_1^L\right) < (1-\phi)\phi ry\left(\sigma - xI_1^L\right)$. Consequently, if $\theta + \phi > 1$, then the HWM may make type $G$ better off as long as

$$\left[(1-\phi)\phi - (1-\theta)\theta\right] ry\left(\sigma - xI_1^L\right) > (\theta - \phi)\, xy\Delta I_0.$$

Using $ry\left(\sigma - xI_1^L\right) = x\Delta I_0 \frac{(1-\pi_0 y)}{(1-\pi_0)\pi_L}$ the above inequality becomes

$$\frac{(1-\phi)\phi - (1-\theta)\theta}{(\theta - \phi)} > \frac{(1-\pi_0)\pi_L}{(1-\pi_0 y)} y.$$

Simplification of the left hand side leads to

$$(\theta + \phi - 1) > \frac{(1-\pi_0)\pi_L}{(1-\pi_0 y)} y.$$

Inserting $\pi_0 = \lambda\theta + (1-\lambda)\phi$, $\pi_L = \lambda_L\theta + (1-\lambda_L)\phi$ and $\lambda_L = \frac{\lambda(1-\theta)}{\lambda(1-\theta)+(1-\lambda)(1-\phi)}$, and further simplifying results in

$$\theta + \phi - 1 > y\theta\phi.$$

asdf. $\diamondsuit$

**Proof of Proposition 8:** In the Outflow Region, with the HWM incumbent investors withdraw funds after a low return until they break even in expectation. Given that incumbent investors break even in the second period, the initial fund size is identical to the case without HWM, $I_0$. The second period fund size following a low return in the first period depends on the amount of the incentive fee paid in the second period. In the case of high returns in the second period, if the net profit, $\sigma - xI_1$, is greater than prorated loss in the first period, $\frac{\widehat{I}_1^L}{I_0(1-x)-\sigma}(\sigma + xI_0)$, then the excess profit, $(\sigma - xI_1) - \frac{\widehat{I}_1^L}{I_0(1-x)-\sigma}(\sigma + xI_0)$, pays incentive fee. Otherwise, no incentive fee will be paid. In either case, the incentive fee paid is less than the case without HWM. Let $q$ stand for the fraction of incentive fee reduction in the second period due to HWM.

Lower incentive fee leads to higher fund size in the second period. Therefore, HWM results in higher fund size following a low return. Let $\Delta I_1$ denote $\widehat{I}_1^L - I_1^L$.The expected additional compensation (in comparison to no HWM) due to higher fund size is therefore

$$(1-\pi_0)x\Delta I_1 > 0$$

However, given that there will be net zero compensation increase, the expected cost which is the expected cost of HWM in terms of lost incentive, has to equal

$$(1-\pi_0)\pi_L y\left(\sigma - xI_1^L\right)q = (1-\pi_0)x\Delta I_1.$$

Rewriting this condition leads to

$$y\left(\sigma - xI_1^L\right)q = \frac{x\Delta I_1}{\pi_L}.$$

Type $G$ and type $B$ do not share these costs and benefits equally. The expected benefit for type $G$ and type $B$ are $(1-\theta)x\Delta I_1$ and $(1-\phi)x\Delta I_1$, respectively. Similarly, the expected cost for type $G$ and type $B$ are $(1-\theta)\theta y\left(\sigma - xI_1^L\right)q$ and $(1-\phi)\phi y\left(\sigma - xI_1^L\right)q$, respectively. Therefore, net benefit for type $G$ is

$$\begin{aligned} & (1-\theta)\left[x\Delta I_1 - \theta y\left(\sigma - xI_1^L\right)q\right] \\ = \; & (1-\theta)\left[x\Delta I_1 - \theta\frac{x\Delta I_1}{\pi_L}\right] \end{aligned}$$

Given that $\theta > \pi_L$ this term is strictly negative. Therefore, type G is strictly worse off due to HWM in the outflow case. . $\Diamond$

**Proof of Proposition 9:**In what follows we will characterize the condition for HWM to increase the efficiency. First, note that both $\Delta I_1$ and $\Delta I_0$ are positive and there is no incentive pay in the second period following a first period low return. Therefore, zero compensation increase due to HWM implies

$$(1-\pi_0)\pi_L y\left(\sigma - xI_1^L\right) = x\left[\Delta I_0(1-\pi_0 y) + (1-\pi_0)\Delta I_1\right]$$

Rewriting this condition leads to

$$\Delta I_1 = \frac{y\left(\sigma - xI_1^L\right)\pi_L}{x} - \frac{\Delta I_0(1-\pi_0 y)}{(1-\pi_0)}$$

Similar to the inflow case, HWM may make type $G$ better off as long as

$$[(1-\phi)\phi-(1-\theta)\theta]\, y\left(\sigma-xI_1^L\right)>x\left[(\theta-\phi)\, y\Delta I_0+(\theta-\phi)\,\Delta I_1\right].$$

Dividing both sides of the above inequality by $(\theta-\phi)$ results in

$$(\theta+\phi-1)\, y\left(\sigma-xI_1^L\right)>xy\Delta I_0+x\Delta I_1.$$

Using zero profit condition $\Delta I_1=\frac{y\left(\sigma-xI_1^L\right)\pi_L}{x}-\frac{\Delta I_0(1-\pi_0 y)}{(1-\pi_0)}$ we have

$$(\theta+\phi-1-\pi_L)\, y\left(\sigma-xI_1^L\right)>x\Delta I_0\left(\frac{y-1}{(1-\pi_0)}\right)$$

Inserting $I_1^L=\frac{\sigma[\pi_L(2-y)-1]}{x(1-y\pi_L)}$, the above inequality becomes

$$(\theta+\phi-1-\pi_L)\, y\frac{2\sigma(1-\pi_L)}{(1-y\pi_L)}>x\Delta I_0\left(\frac{y-1}{(1-\pi_0)}\right) \tag{1}$$

Note that we have $I_0=\frac{\sigma[\pi_0(2-y)-1]}{x(1-y\pi_0)}$ and $\widehat{I}_0=\frac{\sigma[\pi_0(2-y)+(1-\pi_0)(2\pi_L+x-1)-1]}{x[(1-\pi_0 y)+(1-\pi_0)(1-x)]}$ leading to

$$\Delta I_0=2\sigma\left(1-\pi_0\right)\frac{\pi_0(1-x)(1-y)-\pi_L(1-y\pi_0)}{x\left(1-y\pi_0\right)\left(2-x-\pi_0+x\pi_0-y\pi_0\right)}.$$

Inserting the above into ((1)), results in

$$(\theta+\phi-1-\pi_L)\, y\frac{(1-\pi_L)}{(1-y\pi_L)}>\frac{(1-y)\left(\pi_L(1-y\pi_0)-\pi_0(1-x)(1-y)\right)}{(1-y\pi_0)\left((1-\pi_0 y)+(1-x)\left(1-\pi_0\right)\right)} \tag{2}$$

Inserting $\pi_0=(\lambda\theta+(1-\lambda)\phi)$, $\pi_L=(\lambda_L\theta+(1-\lambda_L)\phi)$ and $\lambda_L=\frac{\lambda(1-\theta)}{\lambda(1-\theta)+(1-\lambda)(1-\phi)}$ into (2) leads to the characterization of the condition that is necessary for HWM to increase efficiency.

# References


[1] Anson, M. J. P., 2001, Hedge fund incentive fees and the 'Free Option.' *Journal of Alternative Investments*, Fall, 43-48.

[2] Agarwal, Vikas, and Sugata Ray, 2012, Determinants and implications of fee changes in the hedge fund industry, Working Paper, Georgia State University.

[3] Aragon, George O., and Jun Qian, 2010, The Role of High-water marks in hedge fund compensation, Working Paper, ASU, BC and MIT.

[4] Basak, Suleyman, Anna Pavlova and Alex Shapiro, 2007, Optimal asset allocation and risk shifting in money management, *Review of Financial Studies* 20, 1583-1621.

[5] Berk, Jonathan, and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269-1295.

[6] Berk, Jonathan, and Jules van Binsbergen, 2012, Measuring economic rents in the mutual fund industry, Working Paper, Stanford University.

[7] Brennan, Michael J., 1998, The role of learning in dynamic portfolio decisions, *European Economic Review* 1, 295–306.

[8] Das, Nandita, Richard K. Kish, David L. Muething and Larry W. Taylor, 2002, An overview of the hedge fund industry. Discussion Paper, Lehigh University.

[9] Deuskar, Prachi, Z. Jay Wang, Youchang Wu, and Quoc H. Nguyen, 2012, The Dynamics of Hedge Fund Fees, Working Paper, University of Illinois at Urbana-Champaign.

[10] Drechsler, Itamar, 2011, Risk choice under high-water marks, Working Paper, New York University.

[11] Fung, Bill, David Hseih, Narayan Naik and Tarun Ramadorai, 2008, Hedge funds: Performance, risk and capital formation, *Journal of Finance* 63 (4), 1777-1803.

[12] Gervais, Simon, Anthony Lynch and David K. Musto, 2005, Fund Families as Delegated Monitors of Money Managers. *Review of Financial Studies* 18, 1139-1169.

[13] Goetzmann, William N, Jonathan Ingersoll and Stephen A. Ross, 2003, High-water marks and hedge fund management contracts, *Journal of Finance* 58, 1685-1717

[14] Hodder, James and Jens Carsten Jackwerth, 2007, Incentive contracts and hedge fund management, *Journal of Financial and Quantitative Analysis* 42(4), pp. 811-826.

[15] Huetl, Michael, Otto Loistl and Stephan Zellner, 2008, High-watermark fees, free rides and equal treatment of investors. Working Paper, Vienna University of Economics and Business Administration.

[16] Lee, David K. C., Steven Lwi and Kok Fai Phoon, 2004, An equitable structure for hedge fund incentive fees. *Journal of Investing* (Fall), 31-43.

[17] Naik, Narayan, Tarun Ramadorai and Maria Stromqvist, 2007, Capacity constraints and hedge fund strategy returns. *European Financial Management* 13, 239-256.

[18] Panageas, Stavros, and Mark M. Westerfield, 2007, High water marks: High risk appetites? Hedge fund compensation and portfolio choice, *Journal of Finance* 64(1), 1-36.

[19] Ray, Sugata, and Indraneel Chakraborty, 2007, Effort, risk and walkaway under high water mark style contracts, Working Paper, Wharton School.

[20] Rouah, Fabrice, 2005, Competing Risks for Hedge Fund Survival, Working Paper, McGill University.

[21] Technical Committee of the International Organization of Securities Commissions, 2004, Final report on elements of international regulatory standards on fees and expenses of investment funds. Available at www.sec.gov.

[22] Xia, Yihong, 2001, Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation, *Journal of Finance* 56, 205-246.

**Table I**

Comparative Statics of High-Water Marks

This table provides the net effects of the HWM on assets, returns, and flows in the initial period and after high and low returns. A positive (negative) sign indicates that the presence of a HWM increases (decreases) the equilibrium value of assets, returns, and flows above what would be expected without a HWM. A zero indicates the HWM has no affect on equilibrium outcomes when compared to the case without a HWM.

| Region | Initial Assets | Initial E[R] | Assets After Low | E[R] After Low | Net Flows After Low | Net Flows After High |
|---|---|---|---|---|---|---|
| Inflow | + | - | 0 | + | - | - |
| No-Flow | + | - | + | + | +/- | - |
| Outflow | 0 | 0 | + | 0 | + | 0 |

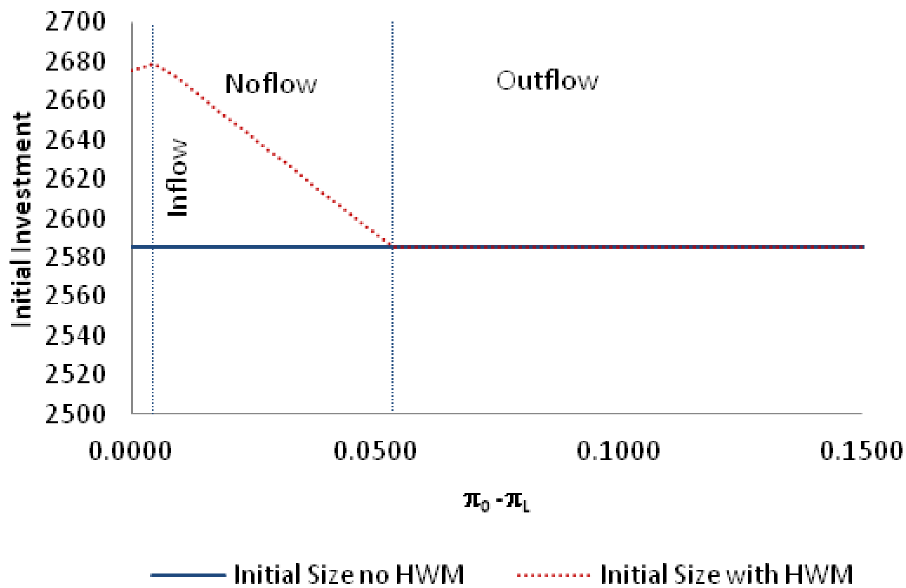


**Figure 1.** Optimal Initial Asset Size. This picture plots the optimal initial asset size of a fund with and without a high-water mark. The x-axis $\pi_0 - \pi_L$ can be interpreted as a measure of precision with larger values corresponding to greater imprecision. In this example, $\pi_0$ is fixed and measures the probability of a good return in the first period while $\pi_L$ updates according to Bayes Rule and is the probability of a good outcome in the second period after observing bad performance in the first period. From the model, this implies that $\lambda$ ranges from 0.66 to 1 and $\theta$ takes a corresponding value such that $\pi_0$ is always 0.797. The remaining parameters are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.

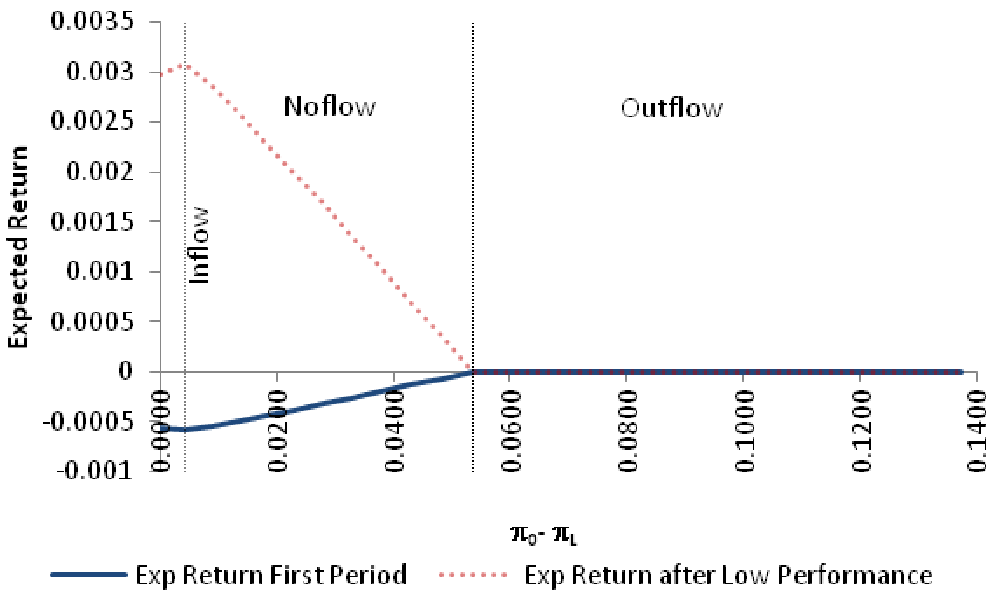


**Figure 2.** Expected returns in the first period and after poor performance for a fund with a high-water mark. The x-axis $\pi_0 - \pi_L$ can be interpreted as a measure of precision with larger values corresponding to greater imprecision. In this example, $\pi_0$ is fixed and measures the probability of a good return in the first period while $\pi_L$ updates according to Bayes Rule and is the probability of a good outcome in the second period after observing bad performance in the first period. From the model, this implies that $\lambda$ ranges from 0.66 to 1 and $\theta$ takes a corresponding value such that $\pi_0$ is always 0.797. The remaining parameters are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.

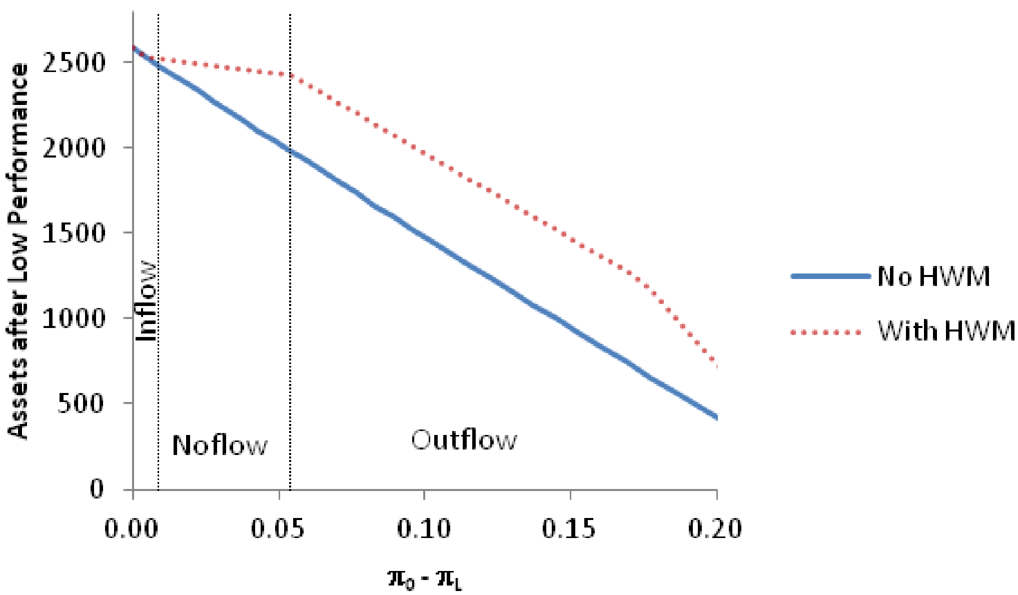


**Figure 3.** Optimal asset size after poor performance for a fund with and without high-water marks. The x-axis $\pi_0 - \pi_L$ can be interpreted as a measure of precision with larger values corresponding to greater imprecision. In this example, $\pi_0$ is fixed and measures the probability of a good return in the first period while $\pi_L$ updates according to Bayes Rule and is the probability of a good outcome in the second period after observing bad performance in the first period. From the model, this implies that $\lambda$ ranges from 0.66 to 1 and $\theta$ takes a corresponding value such that $\pi_0$ is always 0.797. The remaining parameters are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.

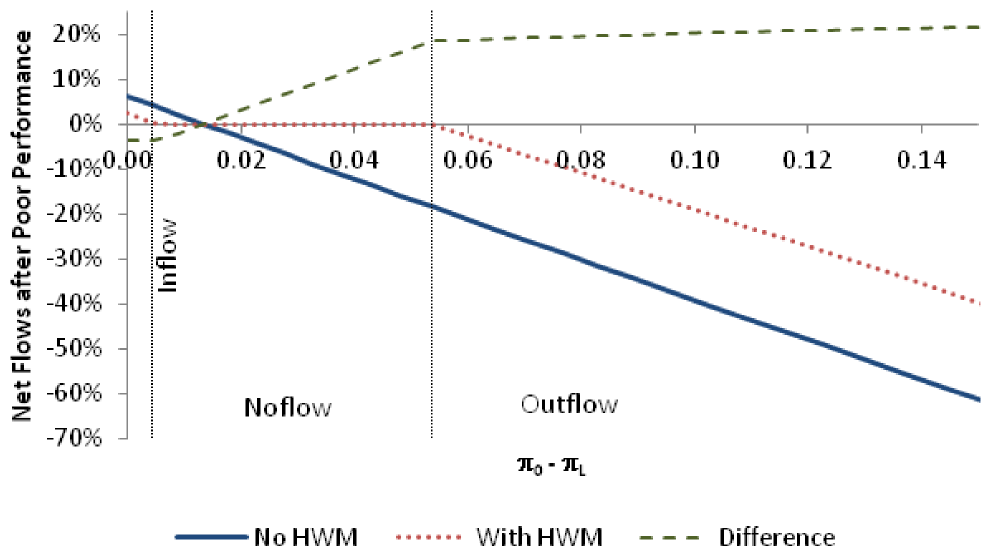


**Figure 4.** Net flows after poor performance for funds with and without high-water marks. Flows are defined as a percent of the assets realized after returns and expenses at the beginning of the second period before new inflows or outflows occur. The x-axis $\pi_0 - \pi_L$ can be interpreted as a measure of precision with larger values corresponding to greater imprecision. In this example, $\pi_0$ is fixed and measures the probability of a good return in the first period while $\pi_L$ updates according to Bayes Rule and is the probability of a good outcome in the second period after observing bad performance in the first period. From the model, this implies that $\lambda$ ranges from 0.66 to 1 and $\theta$ takes a corresponding value such that $\pi_0$ is always 0.797. The remaining parameters are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.

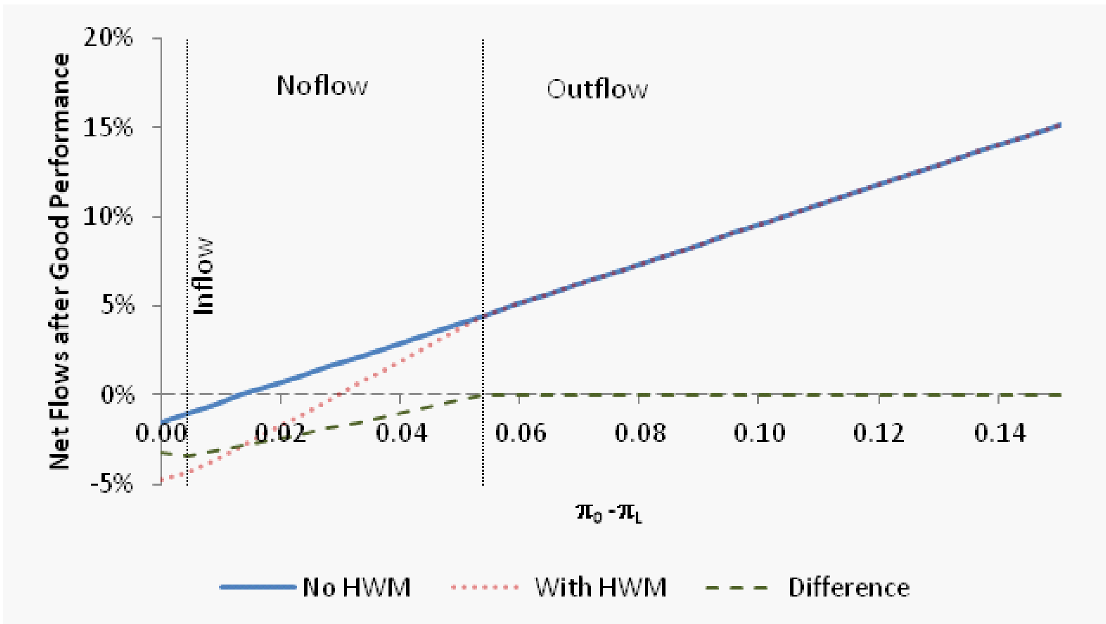


**Figure 5.** Net flows after good performance for funds with and without high-water marks. Flows are defined as a percent of the assets realized after returns and expenses at the beginning of the second period before new inflows or outflows occur. The x-axis $\pi_0 - \pi_L$ can be interpreted as a measure of precision with larger values corresponding to greater imprecision. In this example, $\pi_0$ is fixed and measures the probability of a good return in the first period while $\pi_L$ updates according to Bayes Rule and is the probability of a good outcome in the second period after observing bad performance in the first period. From the model, this implies that $\lambda$ ranges from 0.66 to 1 and $\theta$ takes a corresponding value such that $\pi_0$ is always 0.797. The remaining parameters are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.

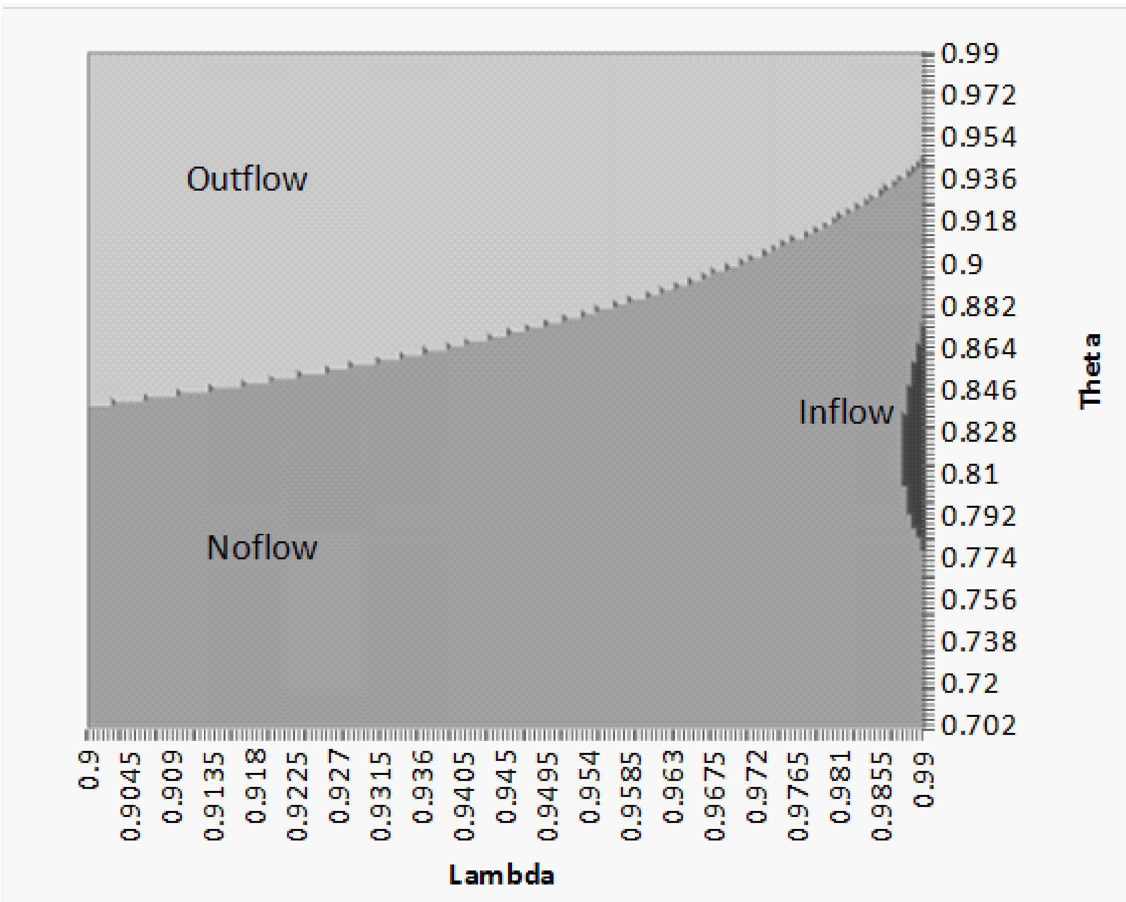


**Figure 6.** Inflow, Outflow, and Noflow regions for different ranges of $\lambda$ and $\theta$. The assumed parameters in the model are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.

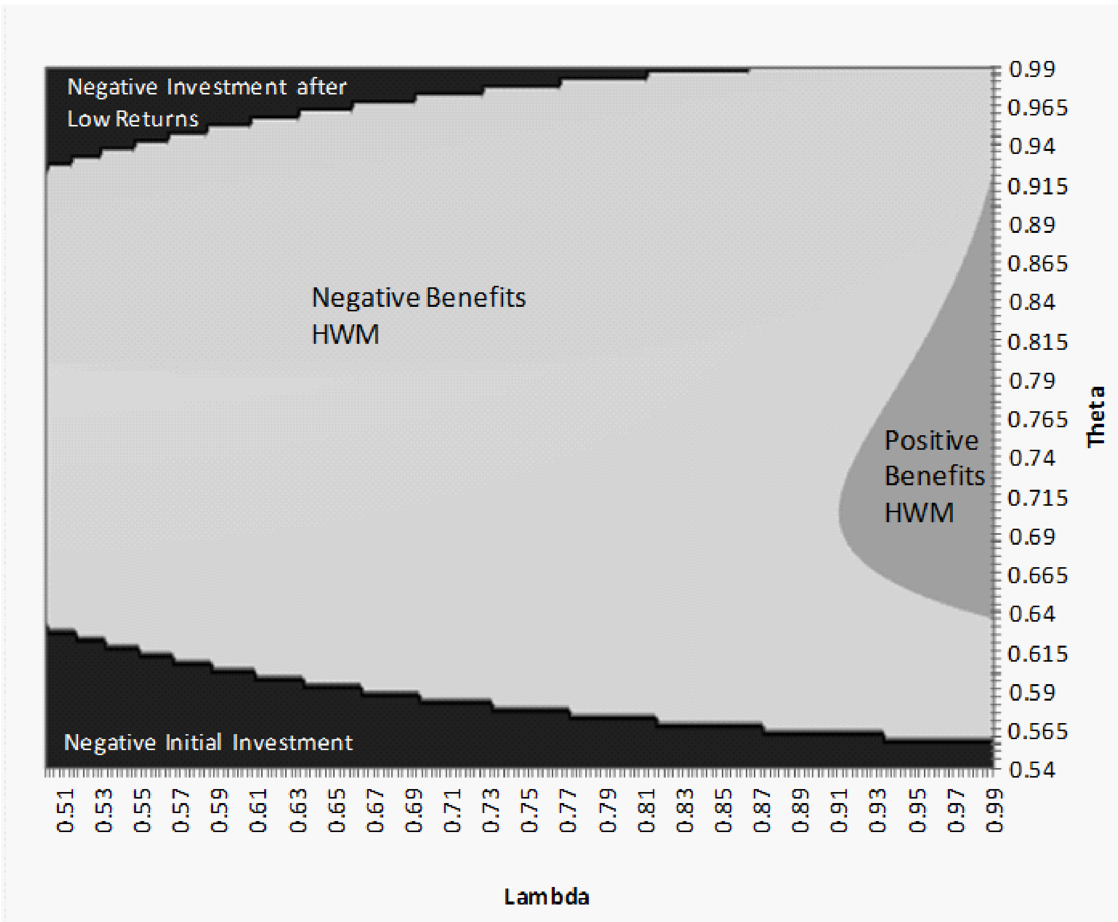


**Figure 7.** Positive and negative benefits of the HWM for different ranges of $\lambda$ and $\theta$. The blackened areas identify regions where the model is not defined because either initial investment is negative or investment after bad returns is negative. The assumed parameters in the model are $\phi = 0.5$, $x = 0.02$, $y = 0.2$,and $\sigma = 100$.