OPTIMAL DYNAMIC MOMENTUM STRATEGIES

KAI LI[†] AND JUN LIU[‡]

[†]Department of Applied Finance

Macquarie University, NSW 2109, Australia

[‡]Rady School of Management

University of California San Diego, La Jolla, California 92093

kai.li@mq.edu.au, junliu@ucsd.edu

Abstract. We explicitly solve for the optimal dynamic trading strategy between a riskless asset and a risky asset with momentum. The optimal portfolio weight depends not only on the momentum, as in Merton's (1971) framework, but also on the historical price path; this contrasts with Merton. Due to their path dependence, optimal portfolio weights have a wide distribution for a given level of momentum; for example, investors may short the risky asset if it has rebound price paths but leverage if it has hump-shaped price paths. This effect tends to be the most significant after large price swings. Path dependence is represented with explicit formulas as well as heuristic statistics.

Key words: Momentum, optimal portfolio, path dependence.

JEL Classification: C32, G11

1. Introduction

The momentum effect has been studied by a tremendous number of papers, but surprisingly, there is little research on optimal momentum strategies. In this paper, we explicitly solve the optimal dynamic portfolio problem when a risky asset has momentum ("momentum asset"). We show that, to optimally exploit momentum, one needs to account for path dependence, as well as momentum. In contrast, the momentum strategies discussed in most papers exploit only momentum.

For an asset with momentum, its optimal portfolio weights depend on a functional of the price path, in addition to momentum. Indeed, a striking feature of momentum is that all the past prices within the look-back period are needed to determine the price dynamics for a finite horizon. Because of this feature, optimal portfolio weights are a functional of these prices. This functional is different for different investment horizons; thus, we need many state variables to construct the optimal portfolios for all horizons. This phenomenon is a consequence of the non-Markovian nature of momentum.

More specifically, we show that the path dependence of an optimal portfolio weight can be captured by a new horizon-dependent variable that differs from momentum. This new variable is a weighted average of past returns, and recent returns receive higher weights than order ones. Indeed, the distribution of the future finite-horizon return (return over a finite horizon) of a momentum asset relies more heavily on recent returns within the look-back period because a recent historical return can be used to predict more future returns than distant ones.

Due to their path dependence, optimal portfolio weights have a wide distribution for a given level of momentum. This distribution becomes more concentrated as the investment horizon decreases, and it converges to an atomic distribution concentrated on the mean-variance portfolio weight as the horizon approaches zero. (We use the term mean-variance portfolio to refer to the strategy used in most of the related empirical studies. The mean-variance portfolio weight equals risk premium divided by return variance and can be justified by a log utility.)

In our paper, path dependence is described through explicit formulas as well as heuristic statistics. The effect of path dependence tends to be the most significant after extreme periods, characterized by substantial price swings. After such periods, the finite-horizon expected returns are large due to momentum; hence, the optimal portfolio weight significantly differs from portfolio weights that ignore path dependence.

More importantly, for a given level of momentum, the distribution of the certainty equivalent wealth (CEW) of the optimal momentum strategy is also wide. It is skewed to the right, indicating that large profits occur occasionally. Indeed, momentum tends to generate large price swings, and the optimal strategy yields higher returns during extreme periods by exploiting these price swings. However, given a level of momentum, the CEW of a suboptimal strategy that ignores path dependence follows an atomic distribution; thus, it is likely to be significantly lower than that of the optimal strategy, which is widely distributed depending on paths.

To illustrate the path dependence of an optimal portfolio, consider two price paths that have the same positive momentum: a rebound path (a path with an early downward trend that later becomes an upward trend) and a hump-shaped path (a path with an early upward trend that later becomes a downward trend). In the case of the rebound path, an investor may short the momentum asset, even if it has positive momentum. Intuitively, each positive historical return implies more positive future returns, and the recent part of a rebound path has more positive returns, leading to a positive expected return over a finite horizon. An investor with a risk aversion coefficient greater than one prefers return reversal over return momentum. Thus, she shorts the asset with positive momentum and effectively constructs a portfolio with return reversal. This suggests that when the asset price has recently experienced a sharp rise, investors should hold fewer assets than the mean-variance portfolio that is always long assets with positive momentum. The optimal portfolio weight for a hump-shaped path tends to follow the opposite pattern.

Path dependence also gives rise to some unique optimal portfolio weight features. For example, the optimal portfolio weight of a momentum asset is differentiable with respect to the horizon only once, and there are many intervals of increases and decreases (horizon bumps). In fact, the functionals of the paths are different for different horizons; hence, the path dependence is also different, leading to horizon bumps.

Our paper extends the classic framework of Merton (1971). Merton assumes that the conditional expected return (and volatility) of an asset is a function of Markovian state variables. In this setting, price dynamics are determined by only these state variables. The optimal portfolio weights and the CEW of all investment horizons also depend only on these variables and are independent of price paths. Because momentum is non-Markovian, Merton's framework cannot be used in this context. Optimal portfolio weights depend on other state variables. As a result, there can be arbitrary differences between Merton's Markovian model and the momentum model in terms of both optimal portfolio weights and the CEW. Additionally, in Merton's framework, the optimal portfolio weight is typically infinitely differentiable and monotonic as a function of the horizon without a horizon bump.

Markov prices are widely assumed in the asset pricing literature, perhaps because of tractability. Efforts have been made to relax this simplifying assumption. Makarov and Rytchkov (2012) show that path-dependent and non-Markovian price dynamics can naturally arise in a rational expectations equilibrium with heterogeneous information. Hommes and Zhu (2014) find that learning tends to generate path dependence if agents do not have perfect knowledge of the relevant economic structure. These results highlight the importance of studying non-Markovian price dynamics. Our paper adds to this body of literature by analytically deriving the optimal dynamic investment strategy. Our solution technique can be used in broader settings where delayed information is useful.

Momentum is one of the most prominent financial market regularities and has been extensively documented. (Jegadeesh and Titman (1993) document momentum in the case of individual U.S. stocks. These results have been extended to stocks in other countries (Rouwenhorst, 1998), industry portfolios (Moskowitz and Grinblatt, 1999), country indices (Asness, Liew and Stevens, 1997), currencies (Okunev and White, 2003), commodities (Gorton, Hayashi and Rouwenhorst, 2013), and exchange-traded futures contracts (Moskowitz, Ooi and Pedersen, 2012; Asness, Moskowitz and Pedersen, 2013).) However, most of the momentum trading strategies mentioned in the existing literature are functions of momentum alone. For example, the buy-and-hold strategy with a six-month look-back period and a six-month holding period studied by Jegadeesh and Titman (1993), among others, sorts stocks based on only momentum. The strategy with a twelve-month look-back period and a one-month holding period used by Moskowitz, Ooi and Pedersen (2012) and Asness, Moskowitz and Pedersen (2013), among others, uses the mean-variance portfolio approach. These strategies are optimal when there is no portfolio re-balancing. Our paper shows that the optimal strategy with re-balancing depends on price paths.

The remainder of paper is organized as follows. Section 2 discusses the momentum model. Section 3 solves the optimal portfolio selection problem. Section 4 studies path dependence, a striking feature of the optimal momentum strategy, Section 5 examines other properties of the optimal strategy, and Section 6 discusses several model extensions. Section 7 concludes this paper while Appendix A presents the main proofs. We discuss return characteristics of momentum assets, some details regarding our calculations and properties of the optimal portfolio weights, and model calibration in the Online Appendices.

2. Model Setup

2.1. **The Momentum Model.** Momentum refers to the tendency of an asset with a sequence of high (low) recent past returns to continue yielding high (low) returns in the near future. To capture this phenomenon, we study a simple model of momentum. More specifically, we assume that the price S_t of a risky momentum asset at time t is given by

$$\frac{dS_t}{S_t} = \big[\alpha m_t + (1-\alpha)\mu + r\big]dt + \sigma dB_t, \tag{2.1}$$

where m_t is momentum, r is the short rate (which is assumed to be a constant), μ is a constant and will be identified later as the unconditional risk premium, α measures the level of momentum in the expected return, σ is a constant, and B_t is a standard Brownian motion.

Momentum is modelled as an equally weighted moving average (MA) of historical excess returns over the following past time interval:

$$m_t = \frac{1}{\tau}\int_{t-\tau}^{t}\Big(\frac{dS_u}{S_u} - r du\Big). \tag{2.2}$$

Following the momentum literature, we call the interval $[t-\tau, t]$ the "look-back period" for momentum.

Model (2.1)–(2.2) is a simple specification of momentum motivated by the empirical literature. In most of the related empirical studies, momentum is modelled as a MA of an asset's historical returns. For example, Moskowitz et al. (2012) document the time series momentum of diverse futures and forward contracts by showing that "*the past 12-month excess return of each instrument is a positive predictor of its future return.*" This evidence has been extended to more asset classes and longer sample periods in, e.g., Georgopoulou and Wang (2017), Hurst, Ooi and Pedersen (2017), and Goyal and Jegadeesh (2018). We acknowledge other approaches to modelling momentum and discuss alternative models in Sections 4.4 and 6.4.

Our model captures only short-run momentum. In real-life data, asset returns also exhibit long-run reversal (e.g., Fama and French, 1988; Poterba and Summers, 1988), which is not considered in our model. Our model could provide a good approximation of the real data-generating process if investors explore only the past one-year returns when making their investment decisions. For example, Moskowitz et al. (2012) find strong short-run (one year) return continuation and weaker long-run reversal, and they use only the past one-year returns to construct a portfolio.

MAs have been widely used in both practice and empirical studies. Because economic conditions change over time, many fund managers use MAs in practice to capture time-varying expected returns (see, e.g., Hurst, Ooi and Pedersen, 2010; 2017; Lo and Hasanhodzic, 2010; Burghardt and Walls, 2011; Narang, 2013). Empirical studies also provide convincing evidence on the predictive power of MAs (see, e.g., Grinblatt and Moskowitz, 2004; Heston and Sadka, 2008; Lewellen, 2015; Brock, Lakonishok and LeBaron, 1992; Neely, Rapach, Tu and Zhou, 2014). In this paper, we study the optimal dynamic trading strategies if MAs can predict asset returns as documented in the literature. Our results can be also used in other settings involving MAs.

We rewrite (2.2) as

$$m_t = \frac{1}{\tau}(\ln S_t - \ln S_{t-\tau}) - r + \frac{\sigma^2}{2}. \tag{2.3}$$

Therefore, momentum also depends on the cumulative return over the look-back period. A key feature of (2.2) is that the average is taken over a moving window $[t-\tau, t]$ with a fixed length τ. Due to this feature, the two prices in (2.3) do not occur simultaneously, which introduce non-Markovian path dependence and mathematical complexity to solving for the optimal portfolio. In this paper, we focus on model (2.1)–(2.2) to elaborate upon path dependence, which is a salient feature of momentum. We will discuss several model extensions in Section 6.

It can be shown that model (2.1)–(2.2) has a unique positive solution for a given positive initial price path over the look-back period, and the corresponding return process is stationary if and only if $-1 < \alpha < 1$. (see Online Appendix III.5.) In this paper, we consider $0 < \alpha < 1$ in order to study momentum.

2.2. **Path Dependence of Finite-Horizon Returns.** The price of the momentum asset in (2.1)–(2.2) is inherently non-Markovian. To define the price process, we need to specify a continuously infinite number of prices S_u, for all u between $-\tau$ and 0 (the look-back period). That is, an infinite number of initial values must be assigned so that the system has an infinite number of dimensions. In contrast, in Merton's (1971) Markovian setting, one needs to specify only the initial values of a finite number of state variables to define the price process (such variables could include price, and predictors of returns and volatility), and the model has a finite number of dimensions.

Due to the non-Markovian dynamics, different historical return paths with the same level of momentum lead to different finite-horizon expected returns and different Sharpe ratios. (Here, finite-horizon expected returns are referred to as the expected returns over a future finite horizon. Different paths with the same level of momentum have the same expected instantaneous return for the next period, which is determined by momentum alone.) Figure 1 illustrates the term structure of the expected returns and Sharpe ratios of three different price paths generated by the momentum model: a rebound path (a path with an early downward trend that later becomes an upward trend), an upward-trending path, and a hump-shaped path (a path with an early upward trend that later becomes a downward trend). (Because the price paths correspond to a diffusion process, the signs of the instantaneous returns change an infinite number of times within any finite interval. The rebound path generally exhibits more negative returns early on and more positive returns later.) All three of the paths have the same beginning and ending prices over the look-back period, and hence they have the same level of positive momentum (the same positive cumulative returns).

With momentum, finite-horizon expected returns tend to be large in magnitude after extreme periods with substantial price swings, such as those seen in rebound and hump-shaped paths. Indeed, each positive historical return implies additional positive future returns; thus, since the recent part of a rebound path has more positive returns, a highly positive return over a finite horizon is expected. Similarly, expected finite-horizon returns tend to be very negative after a hump-shaped path. Expected finite-horizon returns are much smaller in magnitude after an upward-trending path.

Figure 1 also implies that, given the level of momentum, the expected return over a given horizon is widely distributed. In contrast, if the return process is Markovian (e.g., Merton, 1971), the expected return over a given horizon is a single value (rather than a wide distribution of values) for all paths. Path dependence of the finite-horizon returns of momentum assets is found in real data. For example, Da, Gurun and Warachka (2014) show that price paths contain more information than momentum alone, and momentum strategies with finite horizons (e.g., six months or three years) can be significantly improved by exploring the information discreteness exhibited by paths; this suggests that returns are non-Markovian. More return characteristics that are specific to momentum assets and not present in Markov prices are discussed in Online Appendix I.

Model (2.1)–(2.2) is in continuous time, which leads to infinite dimensional price dynamics of the momentum asset. In a discrete-time setting where momentum is measured by the MA of a finite number of past returns over the look-back period, although the corresponding price dynamics are finite dimensional, they still exhibit the path dependence as documented in this paper. In this case, momentum is also not a sufficient statistic of the distribution of finite-horizon returns as observed in model (2.1)–(2.2).

3. The Optimal Dynamic Momentum Strategy

We study the optimal dynamic trading strategy for an investor with an expected utility over terminal wealth W_T at time T and a constant relative risk aversion (CRRA) coefficient $\gamma > 0$. The optimization problem for this investor is given by

$$\sup_{\{\phi_t\}_{t\in[0,T]}} \mathbb{E}_0\left[\frac{W_T^{1-\gamma}}{1-\gamma}\right], \tag{3.1}$$

where ϕ_t is the portfolio weight of the risky momentum asset at time t.

Because there is no Ito's formula for price processes with time delays, the standard dynamic programming approach of Merton (1971) cannot be applied to solve for the optimal trading strategy. (Ito's formula applies to functions of the current values of some stochastic processes but not to the functionals of the paths of processes like ours. The latter can only be addressed with the functional Ito calculus developed recently, e.g., Cont and Fournié (2013) and the references therein.) To the best of our knowledge, there is no known example of closed-form solutions for such systems. (Indeed, most optimal portfolio choice problems cannot be solved in closed form, see, e.g., Haugh, Kogan and Wang, 2006.)

To deal with the path dependence, we derive the solution piecewise. We also use the martingale approach studied in, e.g., Karatzas, Lehoczky and Shreve (1987), Cox and Huang (1989, 1991), and Karatzas and Shreve (1998), which can be applied to non-Markovian prices. We solve the optimization problem in closed form, which provides clean characterizations of the optimal portfolio weight.

In our setting, because the market is complete, the unique state price density is given by

$$\pi_t = \exp\Big\{-\int_0^t r du - \frac{1}{2}\int_0^t \theta_u^2 du - \int_0^t \theta_u dB_u\Big\}, \tag{3.2}$$

where θ_t is the market price of risk, which is given by

$$\theta_t = \frac{\alpha m_t + (1-\alpha)\mu}{\sigma}. \tag{3.3}$$

Because θ_t is path dependent, the price of a dollar at time t in each state is affected by the entire historical return path over the look-back period $[t-\tau, t]$. The standard martingale approach leads to $W_T^* = (\lambda\pi_T)^{-1/\gamma}$, where λ is a Lagrange multiplier. Define

$$\xi_t = \exp\Big\{-\frac{1}{2}\int_0^t \theta_u^2 du - \int_0^t \theta_u dB_u\Big\}, \tag{3.4}$$

which is a martingale. Let $\bar{\xi}_0 = \mathbb{E}_0\big[\xi_T^{(\gamma-1)/\gamma}\big]$. The following lemma provides the optimal wealth and the value function. The corresponding proof is given in Appendix A.1.

Lemma 3.1. *(Optimal wealth.) For an investor with the investment horizon $T-t$ and the constant coefficient of relative risk aversion γ, the optimal wealth process satisfies*

$$W_t^* = W_0 \bar{\xi}_0^{-1} \pi_t^{-1} \mathbb{E}_t\big[\xi_T^{(\gamma-1)/\gamma}\big], \tag{3.5}$$

and the value function satisfies $J = \frac{W_0^{1-\gamma}}{1-\gamma} \bar{\xi}_0^{\gamma} e^{(1-\gamma)rT}$.

Given the optimal wealth W_t^*, the existence of the optimal portfolio is established in the same way as it is in the work of, e.g., Cox and Huang (1989).

Lemma 3.2. *(Optimal portfolio.) The optimal portfolio weight of the risky momentum asset at time t is given by*

$$\phi_t^* = \frac{\alpha m_t + (1-\alpha)\mu}{\sigma^2} + \frac{\psi_t}{\sigma \pi_t W_t^*}, \tag{3.6}$$

where the stochastic process ψ_t that is adapted to the filtration generated by B_t is governed by

$$\pi_t W_t^* = W_0 + \int_0^t \psi_u dB_u. \tag{3.7}$$

The remainder, namely, $1-\phi_t^*$, is invested in the riskless asset.

To derive the optimal portfolio weight, we need to compute $\mathbb{E}_t[\xi_T^{(\gamma-1)/\gamma}]$ in (3.5), which depends on the distribution of the future momentum. However, future momentum is not a function of current momentum; thus, in addition to momentum, we need more (horizon-dependent) state variables that contain additional information about future returns to span the optimal wealth and portfolio.

3.1. **Horizons Shorter Than or Equal To the Look-Back Period.** We first present our results for a case in which the investment horizon is shorter than or equal to the look-back period $0 \le T-t \le \tau$. On the one hand, analyzing this case allows us to keep our proof as short as possible. On the other hand, even in this case, the optimal momentum portfolio can feature the path dependence, which is a salient feature of momentum and the focus of this paper. As shown in Online Appendix I, the path dependence of future returns over long horizons exhibits the same patterns as those over short horizons, and the marginal effect of future "momentum cycles" on path dependence diminishes quickly to zero as the length of a horizon increases. We study the general case covering all horizons in Section 6.3.

Proposition 3.3. *For $0 \leq T-t \leq \tau$, the optimal portfolio weight of the risky momentum asset at time t is given by $\phi_t^* = \phi_t^M + \phi_t^H$, where*

$$\begin{aligned} \phi_t^M &= \frac{\alpha}{\gamma\sigma^2}\int_{t-\tau}^{t}\frac{1}{\tau}\Big(\frac{dS_v}{S_v} - rdv\Big) + \frac{(1-\alpha)\mu}{\gamma\sigma^2}, \\ \phi_t^H &= \frac{(1-\gamma)\alpha}{\gamma\sigma^2}\int_{t-\tau}^{t}\omega_v\Big(\frac{dS_v}{S_v} - rdv\Big) + C_1, \end{aligned} \tag{3.8}$$

ω_v is a deterministic function of v, and C_1 is a constant given by (A.14), *which is shown in Appendix A.2. In addition, ω_v is positive; it increases with v within $v \in [t-\tau, T-\tau]$ but is a constant within $v \in [T-\tau, t]$.*

Proposition 3.3 shows that the optimal portfolio weight consists of two components, as it does in the Markovian setting of Merton (1971). The first component ϕ_t^M is the (myopic) mean-variance portfolio weight. It depends on momentum, which is intuitive because the expected return of the momentum asset depends linearly on momentum. The mean-variance portfolio is optimal for very short horizons and is used in most of the related empirical studies. It is positive if momentum is positive.

The second component, ϕ_t^H, is the product of intertemporal hedging. Unlike the mean-variance portfolio weight, which depends on momentum (an equally weighted average of historical returns over the look-back period), ϕ_t^H depends on a new variable that is an (unequally) weighted average of historical returns. In Merton's framework, the optimal portfolio depends on the state variables that characterize the conditional mean and variance of the asset returns. In contrast, when asset prices have momentum, the historical price paths of the asset provide more information about its future returns than its momentum does. Therefore, the optimal portfolio weight in such a situation depends not only on momentum, which is the conditional mean, but also on other state variables that are functionals of the price paths.

Most studies rely on momentum alone to construct portfolios. Proposition 3.3 shows that this is not optimal in a multiple-period setting if the examined investor rebalances her portfolio. The weighted average of historical instantaneous returns in ϕ_t^H gives rise to the dependence of the optimal portfolio weight on historical return paths. Different paths, even those with the same level of momentum, lead to different portfolio weights. This strategy significantly differs from that in Merton's framework, where the optimal portfolio weight depends only on the current values of the state variables of the expected return and volatility and not on historical paths.

Proposition 3.3 shows that the weight ω_v placed by ϕ_t^H on the historical instantaneous return dS_v/S_v is higher for more recent historical returns. This is because the more recent returns of the momentum asset predict more future returns in a finite horizon. More specifically, consider two past returns, namely, dS_v/S_v and $dS_{v+\Delta}/S_{v+\Delta}$, over the look-back period ($t - \tau \leq v < v + \Delta \leq t$). Both affect the distribution of the future momentum, which determines the value function and optimal portfolio (see Lemmas 3.1 and 3.2). In addition, both returns appear simultaneously in the distribution, and their relative weights geometrically increase with Δ because of momentum. Therefore, the weights placed by ϕ_t^H on these two past returns increase with time.

Figure 2 illustrates the weights (ω_v) placed by ϕ_t^H on the asset's historical instantaneous returns. When the investment horizon is shorter than the look-back period ($T - t < \tau$), the left panel shows that ϕ_t^H weights the recent historical returns in $[t-\tau, T-\tau]$ more heavily and places the same high weight on all the returns in $[T-\tau, t]$. In this case, ϕ_t^H depends on the price path over $[t-\tau, T-\tau]$ and the cumulative return (rather than price path) over $[T-\tau, t]$. When the investment horizon is equal to the look-back period ($T - t = \tau$), the right panel shows that recent historical returns receive higher weights over the entire look-back period $[t - \tau, t]$.

For a given investment horizon, the optimal momentum portfolio weight depends on a finite number of state variables, but for different horizons, these variables are different as illustrated in Figure 2. Since the horizon changes as time changes, we need a large number of state variables to construct the optimal portfolios for all horizons. (Indeed, we need an infinite number of state variables in the continuous-time setting.) This property highlights the non-Markovian nature of momentum. In contrast, in Merton's framework, the optimal portfolio weights for all horizons (including the mean-variance portfolio weight) depend on the same set of state variables, and the number of these variables is finite.

Proposition 3.3 also shows that, for an investor with a risk aversion coefficient greater than one ($\gamma > 1$), ϕ_t^H is negatively affected by the weighted average of historical returns. In addition, the level of ϕ_t^H increases with the level α of momentum (see Figure 2) and decreases with return volatility σ. However, the effects of horizon $T - t$ on ϕ_t^H depend on the entire return path over the look-back period.

In summary, we show that, in addition to momentum, optimal portfolio weights also depend on a weighted average of the instantaneous returns over the look-back period, leading to path dependence. In Sections 4–5, we explore the properties of an optimal portfolio, discussing the novel features of the portfolio weights that are not present in Merton's framework. Online Appendices III.7–III.8 further discuss some of the limiting properties of the optimal portfolio weights with simplified expressions and the portfolio weights in terms of cumulative returns, providing more insight into these weights.

4. Path Dependence

In this section, we study the path dependence of both the optimal portfolio weight and portfolio performance.

4.1. **Path Dependence of the Optimal Portfolio Weight.** Proposition 3.3 shows that an optimal portfolio is a functional of the price path over the look-back period. This feature is closely related to the path dependence of the portfolio weight. In Merton's framework, the optimal portfolio weight is completely determined by the state variables that specify the expected return and volatility. However, for the momentum asset, even though momentum specifies the expected return, it does not completely determine the optimal portfolio weight because price paths can predict future finite-horizon returns. Figure 3 plots the distributions of the examined optimal portfolio weights. To highlight the effect of paths, we choose a fixed value for momentum $m_0 = \mu$, i.e., its long-run mean. The results are based on 10,000 simulated historical paths generated by model (2.1)–(2.2).

Optimal portfolio weights over a one-year investment horizon vary widely between -0.2 and 0.5. This result is intuitive. Different paths with the same level of momentum lead to different distributions of future returns over an investment horizon, as shown in Section 2.2. Therefore, the optimal momentum portfolio weights can be significantly different for different paths. This wide distribution of portfolio weights leads to large deviations from the results of Merton (1971). For example, Figure 3 illustrates the optimal portfolio weights $\widehat{\phi}^1$ for a Markovian approximation model specified shortly in Section 4.4. Given an investment horizon, $\widehat{\phi}^1$ is the same for all the paths. It is close to the mean-variance portfolio weight but can even be qualitatively different from the optimal momentum portfolio weight for certain paths. Indeed, the instantaneous return and finite-horizon return of the asset are characterized by different sets of state variables in the momentum model but the same set of state variables in a Markovian price model.

Because the mean-variance portfolio weights depend on momentum alone, they are the same for different paths and different investment horizons. Figure 3 shows that the distribution of the optimal portfolio weights becomes more concentrated for shorter horizons. The optimal portfolio weights converge to the mean-variance weight as the horizon approaches zero. Because the level of momentum is positive for all the paths, the average optimal portfolio weight is lower than the mean-variance weight for $\gamma > 1$.

To shed more light on the path dependence of optimal portfolio weights, Figure 4 compares the portfolio weight dynamics of the three historical price paths illustrated in Figure 1. The portfolio weights are plotted against the investment horizons in the left, middle, and right panels for the rebound, upward-trending, and hump-shaped paths, respectively.

The mean-variance portfolio weights for the three paths are the same because all the paths have the same level of momentum ($m_0 = \mu > 0$). However, the optimal portfolio weights differ for these three paths and depend on the investment horizons. The optimal portfolio weight for the rebound path, which is illustrated in the left panel of Figure 4, is lower than the mean-variance weight, indicating a negative hedging demand. For long investment horizons, the portfolio weight can be negative ($T - t > 0.5$) even though the level of momentum is positive. Indeed, each positive historical return implies more positive future returns; since the recent part of a rebound path exhibits more positive returns, the expected return over a finite horizon is highly positive. Accordingly, an investor with $\gamma > 1$ holds fewer risky assets. In other words, the left panel suggests that investors should hold fewer stocks than indicated by the mean-variance strategy when recent prices rose to the current price. The right panel shows that the portfolio strategy for an asset with a hump-shaped path tends to follow the opposite pattern. This suggests that an investor should hold more stocks than indicated by the mean-variance strategy when the price has recently experienced a dramatic decline.

The short position illustrated in the left panel can also be understood as follows. Momentum tends to generate return persistence. For example, He and Li (2015) show that momentum tends to destabilize the price process by leading to large price swings. However, an investor with a risk aversion coefficient greater than one prefers a reversal in returns to momentum, reflecting an attempt to minimize the (unanticipated) variability in wealth over time (Merton, 1971). Accordingly, such an investor shorts assets with positive momentum and effectively constructs a portfolio with return reversal. This also causes her wealth process to exhibit fewer swings than the price process of the momentum asset.

More importantly, Figure 4 shows that different price paths can qualitatively change the optimal portfolio strategy. The optimal portfolio weights of assets with rebound paths and hump-shaped paths that have the same level of momentum can even have different signs. This further implies that the weighted average of the historical returns in the optimal portfolio weight, which characterizes the hedging demand, shown in Proposition 3.3 can be quite large in magnitude for extreme periods with large price swings. Figure 4 also shows that the left and right tails of the portfolio weight distribution in Figure 3 tend to be generated by the rebound and hump-shaped paths, respectively.

Because the expected finite-horizon returns after an upward-trending path are much smaller in magnitude than those after extreme periods as illustrated in Figure 1, the optimal portfolio weight shown in the middle panel of Figure 4 has fewer deviations from the mean-variance portfolio weight than the weights associated with the rebound and hump-shaped paths.

4.2. **Heuristic Statistic of Path Dependence.** The previous subsection shows that the optimal portfolio weight significantly differs from the mean-variance weight, and this difference tends to be the largest after extreme periods with large price swings, such as those exhibited by rebound paths or hump-shaped paths. To capture the path dependence of this weight, we provide a heuristic statistic in this subsection, which is defined as

$$X_{max} = \max_{t',t''\in[t-\tau,\ t]} \left\{(\ln S_{t'} - \ln S_{t''})^2\right\}. \tag{4.1}$$

This statistic is the maximum of the squared historical cumulative returns over the look-back period. It measures the magnitude of the price swings over the look-back period and is path-dependent. A higher X_{max} value indicates that the given asset price is more likely to exhibit large swings.

Figure 5 plots the optimal portfolio weight ϕ^* against X_{max}. It shows that the deviation of the optimal portfolio weight from the mean-variance weight is the largest during extreme periods with large price swings as measured by X_{max}. (We can rewrite (4.1) as $X_{max} = \left(\ln \frac{S_t}{\max_{u\in[t-\tau,\ t]}\{S_u\}} - \ln \frac{S_u}{\min_{u\in[t-\tau,\ t]}\{S_t\}}\right)^2$, where, for $\tau = 1$ year, $\frac{S_t}{\max_{u\in[t-\tau,\ t]}\{S_u\}}$ measures nearness to the 52-week high, which is found to proxy for underreaction, e.g., George and Hwang (2004) and Li and Yu (2012), and $\frac{S_t}{\min_{u\in[t-\tau,\ t]}\{S_u\}}$ is nearness to the 52-week low.)

4.3. **Economic Value of Path Dependence.** To assess the economic value of path dependence, we compare the optimal momentum strategy to the mean-variance strategy that ignores path dependence in terms of certainty equivalents. We define the certainty equivalent wealth (CEW) of a strategy for an investor as the amount of wealth that makes the investor indifferent between receiving the CEW with certainty at the terminal time T and having \$1 today to invest until time T using the strategy. Define the present value of CEW as $R = e^{-rT}\text{CEW}$. Therefore, R represents the expected gross return over the riskless return delivered by the strategy. We compute this in Online Appendix III.1.

Given momentum ($m_0 = \mu$), Figure 6 illustrates the present value of the CEW of the optimal momentum strategy (R^*) and that of the CEW of the mean-variance strategy (R^M). The results are based on 10,000 simulated historical paths generated from model (2.1)–(2.2). We set $\gamma = 2$ throughout the paper.

The left panel of Figure 6 plots the distributions of the present value of the CEW. The distribution for the optimal momentum strategy (R^*) is wide, demonstrating that the performance of the optimal strategy is crucially dependent upon the price paths. Therefore, the performance of the optimal strategy deviates considerably from that of the strategy of Merton, in which, given the value of the state variable (momentum), the CEW should be a single value (an atomic distribution). Because the optimal portfolio weight depends on different state variables for different investment horizons, R^* exploits these a large number of state variables when employing the strategy until time T. The left panel shows that the CEW of the mean-variance strategy (R^M) also depends on paths. This is due to the path dependence of the momentum asset's price dynamics.

As shown in the left panel of Figure 6, R^* is greater than 1, as expected. R^M is also greater than 1, showing that the performance of the mean-variance portfolio is better than that of the riskless asset for all the price paths, as proved by (III.2) in Online Appendix III.1. Intuitively, the mean-variance strategy always leads to positive risk premiums. The mean values of R^* and R^M are 1.023 and 1.008, respectively, showing that exploiting paths triples the average expected returns of the mean-variance strategy. The CEW ratios R^*/R^M vary between 1.01 and 1.07; thus, accounting for paths is expected to increase returns by up to 7% per year. Moreover, the standard deviation of R^* is 0.01; this is much smaller than the return volatility of the momentum asset, reflecting the return "smoothing" of the optimal strategy with $\gamma > 1$.

The right panel of Figure 6 plots the CEW ratio against X_{max}, which is defined in (4.1) and measures the magnitude of the price swings over the look-back period. It shows that the outperformance of the optimal strategy over the mean-variance strategy tends to be the highest after extreme periods with substantial price swings. For example, the present values of the CEW of the optimal strategy after the rebound and hump-shaped paths illustrated in Figure 1 are 1.07 and 1.03, respectively; however, after an upward-trending path with the same level of momentum, the CEW is only 1.02. Indeed, momentum tends to generate large price swings, and the optimal strategy yields higher returns by exploiting these swings. As a result, the optimal strategy occasionally delivers large profits during these extreme periods with large swings, leading the CEW distribution for the optimal strategy to skew to the right (with skewness of 2.61).

Put differently, extreme periods with large price swings, such as those exhibited by rebound and hump-shaped paths, tend to generate higher absolute values of Sharpe ratios than upward-trending paths with the same level of momentum, as shown in Section 2.2. Therefore, the CEW that depends on the sum of squared Sharpe ratios (the market price of risk) over the investment horizon tends to be highest after these extreme periods.

In summary, we show that historical price paths can be used to improve the profitability of momentum strategies, especially in the case of extreme periods with large price swings.

4.4. **Historical Average Approximation.** The MA (2.2) method, though it faithfully follows the literature and is empirically robust to implement, is non-Markovian. Because some types of historical averages (HAs) are Markovian and thus more tractable than MAs, it is sometimes suggested that HAs are used to approximate MAs. One such approximation could be:

$$\begin{aligned} \frac{dS_t}{S_t} &= \Big[\sum_{i=1}^{N} \alpha_i m_{it} + \Big(1 - \sum_{i=1}^{N} \alpha_i\Big)\mu + r\Big] dt + \sigma dB_t, \\ m_{it} &= \int_{-\infty}^{t} \frac{1}{\tau_i} e^{-(t-u)/\tau_i} \Big(\frac{dS_u}{S_u} - r du\Big), \end{aligned} \tag{4.2}$$

where α_i and τ_i are positive constants. The HA variable, m_{it}, follows a mean-reverting process, namely, $dm_{it} = \frac{1}{\tau_i}[(1 - \sum_{j=1}^{N} \alpha_j)\mu - (m_{it} - \sum_{j=1}^{N} \alpha_j m_{jt})]dt + \frac{\sigma}{\tau_i} dB_t$. This equation shows that an HA with a high decay rate ($1/\tau_i$) has a high level of volatility.

The price dynamics (4.2) are Markovian, and the corresponding portfolio selection problem is within Merton's framework. Asset price processes with mean-reverting expected returns such as those in (4.2) have been extensively studied, e.g., by Kim and Omberg (1996), Liu (2007), and Koijen, Rodríguez and Sbuelz (2009), among others. The following proposition summarizes the optimal portfolio weight of a risky asset according to model (4.2).

Proposition 4.1. *The optimal portfolio weight of a risky asset with a price of* (4.2) *with N Markovian state variables is given by*

$$\widehat{\phi}_t^N = \frac{1}{\gamma\sigma^2}\Big[\sum_{i=1}^{N} \alpha_i m_{it} + \Big(1 - \sum_{i=1}^{N} \alpha_i\Big)\mu\Big] + \sum_{i=1}^{N} \frac{1}{\tau_i}\Big(\sum_{j=1}^{N} A_{ijt} m_{jt} + A_{it}\Big), \tag{4.3}$$

where A_{ijt} and A_{it} are deterministic functions given by (II.1) *in Online Appendix II.1.*

Proposition 4.1 shows that the optimal HA portfolio weight depends on only m_i and is independent of historical price paths. Given a positive level of momentum, the hedging demand is *always* negative (positive) for $\gamma > 1$ ($\gamma < 1$).

We calibrate model (4.2) based on the size factor using the maximum likelihood method. Estimation details are presented in Online Appendix II.3. We focus on the cases of $N = 1, 2$. For both cases, coefficients of α_1 and α_2 are positive, and the t-statistics of them are lower than that for the MA model (2.2).

We compare the performance of the optimal momentum strategy and the optimal HA s-trategies. By applying the HA strategies (4.3) to the momentum asset (2.1), the left panel of Figure 8 illustrates the distributions of the present values of the resulting CEW, $\widehat{R}^N$, $N = 1, 2$. (See Online Appendix II.2 for the calculation details regarding the CEW of the HA strategies.) While the CEW of both the optimal momentum strategy and the mean-variance strategy is greater than 1, as shown in Figure 6, the $\widehat{R}^N$ resulting from the HA approximations can be less than 1, which means that the HA strategies underperform even the riskless asset for certain paths. The mean value of $\widehat{R}^1$ is 1.001, which is lower than the mean value of R^*, namely 1.023, suggesting that exploiting paths increases the expected returns by 2% per year. We find that including more Markovian state variables (m_{1t}, m_{2t}) increases the mean value of CEW ($\widehat{R}^2 = 1.004$). In addition, the standard deviations of the CEW of the optimal momentum strategy, the HA strategy with $N = 1$, and the HA strategy with $N = 2$ are 0.65%, 0.78%, and 0.74%, respectively, the skewness are 2.61, -1.73, and -1.18, respectively, and the kurtosis are 14.07, 8.03, and 7.08, respectively. Thus, more Markovian state variables tend to increase the expected returns and skewness of the HA strategies and decrease the return volatility and kurtosis. Intuitively, the use of a sufficient number of HA variables in (4.2) can enable the expected instantaneous return of the momentum asset in (2.2) to be approximated arbitrarily well, improving the performance of the corresponding HA strategies.

The right panel of Figure 8 illustrates the CEW ratios $R^*/\widehat{R}^N$. We also find that the coefficients from the regressions of the CEW ratios onto X_{max} are highly significantly positive. Thus, the optimal momentum strategy delivers its highest outperformance over the HA strategies during extreme periods with large price swings, implying that the HA approximations make major forecasting errors during such periods.

5. Other Properties of the Optimal Dynamic Momentum Strategy

5.1. **Horizon Dependence.** There are many intervals of increases and decreases in the portfolio weight as a function of horizon (horizon bumps), as illustrated in Figure 4. In contrast, the optimal portfolio weight in Merton's framework is typically infinitely differentiable and monotonic as a function of horizon.

Figure 7 provides insight into these horizon bumps. The solid blue line is the logarithm of a price path during $[t-\tau, t]$ generated from the momentum model, and the dashed red line illustrates the optimal portfolio weights for different investment horizons $T-t \in [0,\tau]$. The horizon bumps depend on the sign of $\frac{\partial \phi_0^*}{\partial T}$. Online Appendix III.6 demonstrates that the portfolio weight increases (decreases) with T when the cumulative return $\ln S_t - \ln S_{T-\tau} + C_6$ is negative (positive), where C_6 is a constant given by (III.37). For example, $\ln S_t - \ln S_{T-\tau} + C_6$ is positive when $T-t < 0.2$, but becomes negative when $0.2 < T-t < 0.5$. Accordingly, Figure 7 shows that the portfolio weight decreases with the horizon when $T-t < 0.2$ while increasing when $0.2 < T-t < 0.5$.

In fact, the optimal portfolio weight for different investment horizons depends on the different state variables employed. Effectively, we need a large number of state variables to constitute a sufficient statistic of portfolio weight if we consider horizon dependence. Changes in the sets of state variables lead to horizon bumps.

Even though portfolio weight as a function of investment horizon exhibits bumps, it is still a differentiable function. Corollary III.2 in Online Appendix III.6 shows that the portfolio weight is differentiable with respect to the horizon only once. In contrast, the optimal portfolio weight for Markovian prices is typically infinitely differentiable and monotonic as a function of horizon.

5.2. **Separating Momentum and Historical Price Paths.** To further our understanding of path dependence, we develop a more general price process (than equation (2.1)), which is given as

$$ds_t = [(\alpha_1 s_t - \alpha_2 s_{t-\tau})/\tau + (1-\alpha)(r+\mu-\sigma^2/2)]dt + \sigma dB_t, \tag{5.1}$$

where $s_t = \ln S_t$ and α_1 and α_2 are parameters. For $\alpha_1 = \alpha_2 = \alpha$, (5.1) reduces to the momentum model (2.1) and (2.2) in Section 2. For $\alpha_2 = 0$, (5.1) becomes a Markovian process.

For $\alpha_1 = 0$, momentum is "turned off"and only path dependence is relevant:

$$ds_t = [-\alpha_2 s_{t-\tau}/\tau + (1-\alpha)(r+\mu-\sigma^2/2)]dt + \sigma dB_t.$$

The corresponding optimal portfolio weight when $T-t \le \tau$ is given by

$$\phi_t^* = \frac{\alpha(r-\sigma^2/2) + (1-\alpha)\mu - \alpha_2 s_{t-\tau}/\tau}{\gamma\sigma^2}.$$

In this case, the hedging demand disappears because the path is uncorrelated with the innovation of asset price. This implies that the new state variable ϕ_t^H in Proposition 3.3 is caused by the joint impact of s_t and $s_{t-\tau}$ on the expected returns.

5.3. **Dependence on Other Variables.** When the price path S_u changes to cS_u for all $u \in [t-\tau, t]$, where $c > 0$ is a constant, ϕ_t^M and ϕ_t^H do not change. This is because momentum is defined in terms of returns; hence, price level does not affect future returns. Thus, both demand components depend on historical returns. This is demonstrated in Online Appendix III.9.

In addition, we find a nonmonotonic dependence of the portfolio weight on the risk aversion coefficient γ. For $\gamma < 1$, there is a finite critical horizon $\sqrt{\gamma}\tau/(2\alpha)\ln[(1+\sqrt{\gamma})/(1-\sqrt{\gamma})]$, at which both the portfolio weight and expected utility approach infinity. This phenomenon, which is referred to as "nirvana", has been observed when the expected return follows an Ornstein-Uhlenbeck process (Kim and Omberg, 1996). Our paper shows that nirvana is also possible in the context of a momentum asset if γ is small enough.

6. Model Extensions

This section discusses several model extensions and presents the optimal portfolio weight for any investment horizons.

6.1. **Cross-Sectional Momentum.** To elaborate upon path dependence, this paper focuses on a case with a single momentum asset in model (2.1)–(2.2). Such a case is termed time series momentum in the existing studies on this topic. In this section, we extend our model to a case involving multiple risky assets, which can be used to study cross-sectional momentum.

6.1.1. *The Model.* We assume that there are K risk factors that are captured by K factor assets. Denote $dR_{k,t}^F = dS_{k,t}^F/S_{k,t}^F$ as the return of factor asset k at time t, where $S_{k,t}^F$ is the asset's price. The factor asset return follows

$$dR_{k,t}^F = \alpha_k dt + dB_{k,t}^F, \tag{6.1}$$

where α_k is a constant and $dB_{k,t}^F$ is a standard Brownian motion that satisfies $dB_{k,t}^F dB_{j,t}^F = 0$ for $k \neq j$. In (6.1), the returns of each factor asset are IID, and the factors are mutually independent.

In addition to the K factor assets, there are N momentum assets. Their returns follow

$$d\mathbf{R}_t^M - rdt = (a + b\mathbf{m}_t)dt + \boldsymbol{\beta} d\mathbf{R}_t^F + \boldsymbol{\Sigma} d\mathbf{B}_t^M, \tag{6.2}$$

where $d\mathbf{R}_t^M = (dS_{1,t}/S_{1,t}, dS_{2,t}/S_{2,t}, \cdots, dS_{N,t}/S_{N,t})'$ is a vector of the returns of the momentum assets, $S_{i,t}$ is momentum asset i's price, a and $b > 0$ are constants, $d\mathbf{R}_t^F = (dR_{1,t}^F, dR_{2,t}^F, \cdots, dR_{N,t}^F)'$, $\boldsymbol{\beta}$ is an $N \times K$ matrix of factor loadings, $\boldsymbol{\Sigma}$ is a volatility matrix, and $d\mathbf{B}_t^M = (dB_{1,t}^M, dB_{2,t}^M, \cdots, dB_{N,t}^M)'$ is a vector of standard Brownian motions. We assume that $\boldsymbol{\Sigma}$ is a diagonal matrix whose iith element is given by σ_i. In this case, the idiosyncratic risks of the momentum assets in (6.2) are captured. (Our analytical results also hold for nondiagonal $\boldsymbol{\Sigma}$.) This case is also studied in the empirical literature, e.g., Grinblatt and Moskowitz (2004), Heston and Sadka (2008), and Lewellen (2015). We assume that all the Brownian motions are mutually independent. The $\mathbf{m}_t = (m_{1,t}, m_{2,t}, \cdots, m_{N,t})'$ is a vector of momentum given by

$$\mathbf{m}_t = \frac{1}{\tau} \int_{t-\tau}^{t} (d\mathbf{R}_u^M - r du). \tag{6.3}$$

In model (6.2), momentum predicts the cross-section of risk-adjusted returns according to certain factors (6.1).

We choose the same slope coefficient, b, for all the momentum assets. This implies cross-sectional momentum and is also a standard assumption in the existing studies on cross-sectional returns (e.g., Fama and MacBeth, 1973). (If b is different for different assets, the model may not generate momentum, and assets with good relative performance may have lower expected risk-adjusted returns than other assets. We can still derive closed-form solutions when different assets have different coefficients.) Models of cross-sectional returns that satisfy (6.2) include those of Grinblatt and Moskowitz (2004), Heston and Sadka (2008), and Lewellen (2015), among others, who show that momentum positively predicts cross-sectional firm returns. In untabulated results, we examine the buy-and-hold momentum strategies proposed by Jegadeesh and Titman (1993) using the data generated from model (6.2), and we find that our model can produce the cross-sectional momentum effect documented in the work of Jegadeesh and Titman (1993). In this sense, we have a realistic and tractable model of cross-sectional momentum.

6.1.2. *The Optimal Cross-Sectional Momentum Strategy.* Now, we study optimal dynamic cross-sectional momentum strategies. The market is complete, and the Cox-Huang approach still applies. The market prices of risk are given by

$$\boldsymbol{\theta}_t = \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}, \quad \text{where} \quad \boldsymbol{\Omega} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{\beta} & \boldsymbol{\Sigma} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\alpha} - r \\ a + b\mathbf{m} \end{pmatrix}, \tag{6.4}$$

$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_K)'$, and $\mathbf{I}$ is a $K \times K$ identical matrix. The market price of risk is jointly determined by all the assets' momentum. Proposition 6.1 summarizes the optimal dynamic cross-sectional momentum strategies that have closed-form solutions. The proof is given in Online Appendix III.2.

Proposition 6.1. *When $0 \leq T \leq \tau$, the optimal portfolio weights of momentum assets are given by $\boldsymbol{\phi}^* = \boldsymbol{\phi}^M + \boldsymbol{\phi}^H$, where*

$$\boldsymbol{\phi}^M = \frac{1}{\gamma}\big[(\boldsymbol{\Sigma}^{-2}+\boldsymbol{\beta}\boldsymbol{\beta}')(a+b\boldsymbol{m}+\boldsymbol{\beta}\boldsymbol{\alpha})-\boldsymbol{\beta}(\boldsymbol{\alpha}-r)\big], \quad \boldsymbol{\phi}^H = (1-\gamma)(\boldsymbol{\Sigma}^{-2}+\boldsymbol{\beta}\boldsymbol{\beta}')\int_{-\tau}^{0}\omega_v(d\boldsymbol{R}_v-rdv)+\boldsymbol{C},$$

ω_v are deterministic weights, and $\boldsymbol{C}$ is a vector of constants given by (III.17) *in Online Appendix III.2. The optimal portfolio weights of factor assets are given by $\boldsymbol{\phi}^{F*} = \frac{1}{\gamma}\big[\boldsymbol{\alpha} - r - \boldsymbol{\beta}'(a + b\boldsymbol{m} + \boldsymbol{\beta}\boldsymbol{\alpha})\big]$, and the remainder, $1 - \boldsymbol{1}'\begin{pmatrix}\boldsymbol{\phi}^* \\ \boldsymbol{\phi}^{F*}\end{pmatrix}$, is invested in a riskless asset.*

Consistent with the results regarding the optimal portfolio weight for the time series momentum model (2.1)–(2.2) developed in Section 2, Proposition 6.1 shows that the portfolio weights of the momentum assets consist of two components. The first component, $\boldsymbol{\phi}^M$, represents the mean-variance portfolio weights. It depends on momentum. In general, an asset that has had a relatively good recent performance receives a higher mean-variance portfolio weight than other assets. The second component, $\boldsymbol{\phi}^H$, represents the intertemporal hedging demands. It depends on new variables, and each variable is a weighted average of the historical returns of a stock over the given look-back period.

The key results regarding path dependence and horizon dependence observed from model (2.1)–(2.2) also apply to optimal dynamic cross-sectional momentum strategies. Due to path dependence, it is optimal to invest more in momentum assets and less using the mean-variance strategy if the return path is hump-shaped; in the case of a rebound path, it is optimal to invest less in momentum assets. As a result, the optimal strategy places more portfolio weight on a winner asset (i.e., with a relatively good recent performance as measured by momentum) whose positive past returns occurred more recently than those of other winners. Indeed, a key feature of both types of momentum is the ability to predict returns with a moving average (MA) of past returns. (Moskowitz et al. (2012) show that the predictability of assets' monthly excess returns by their lagged one-year returns is the main driving force of both types of momentum. Ehsani and Linnainmaa (2021) find that factor returns exhibit time series momentum and time series factor momentum fully subsumes cross-sectional momentum in individual stock returns.) Model (2.1)–(2.2) characterizes this feature, and the insights provided in this paper also apply to the cross-sectional momentum.

Proposition 6.1 shows that an asset's portfolio weight depends not only on its own momentum and price path but also on the momentum and price paths of other assets due to the term $\boldsymbol{\beta}\boldsymbol{\beta}'$, which results from the factor loadings of momentum assets. If there are no common factors driving all the assets' returns, then the optimal portfolio weight of each asset depends on its own past returns alone.

6.2. **Sharpe Ratio Momentum.** While the momentum literature focuses largely on price momentum, some papers also document "Sharpe ratio momentum", where momentum is calculated as the MA of volatility-normalized asset returns, e.g., Rachev, Jašić, Stoyanov and Fabozzi (2007). To capture this effect, we define the normalized return process R_t as $dR_t = \frac{1}{\sigma_t}(\frac{dS_t}{S_t} - r dt)$, where σ_t is the instantaneous volatility of the given asset returns at time t, which can be time-varying (e.g., Kim and Omberg, 1996). In this case, a momentum asset's price follows

$$dR_t = [\alpha m_t^R + (1-\alpha)\mu^R]dt + dB_t, \tag{6.5}$$

where momentum, m_t^R, is defined over the normalized returns following

$$m_t^R = \frac{1}{\tau}\int_{t-\tau}^{t} dR_v. \tag{6.6}$$

In particular, when $\sigma_t \equiv \sigma$ is a constant, model (6.5)–(6.6) reduces to (2.1)–(2.2).

We now study the optimization problem presented in (3.1), and the following proposition summarizes the optimal Sharpe ratio momentum portfolio for this case.

Proposition 6.2. *When $0 \le T \le \tau$, the optimal portfolio weight of a risky momentum asset at time t is given by $\phi_t^{R*} = \phi_t^{RM} + \phi_t^{RH}$, where*

$$\phi_t^{RM} = \frac{1}{\gamma\sigma_t}\Big[\frac{\alpha}{\tau}\int_{t-\tau}^{t} dR_v + (1-\alpha)\mu^R\Big], \quad \phi_t^{RH} = \frac{1-\gamma}{\sigma_t}\Big[\int_{t-\tau}^{t} \omega^R(v,t)dR_v + C^R(t)\Big];$$

$\omega^R(v,t)$ and $C^R(t)$ are deterministic functions given by (III.21) *in Online Appendix III.3. In addition, ω^R is positive; it increases with v when $v \in [t-\tau, T-\tau]$ but is a constant when $v \in [T-\tau, t]$.*

Proposition 6.2 shows that the optimal Sharpe ratio momentum portfolio weight also features path dependence as observed for price momentum. In this case, the weights in the weighted average of the historical instantaneous returns in ϕ_t^{RH} depend also on instantaneous return volatilities.

6.3. **Horizons Longer Than the Look-Back Period.** In this paper, we mainly focus on cases where the investment horizon is shorter than or equal to the look-back period $0 \le T-t \le \tau$ to illustrate path dependence. In this section, we derive the optimal portfolio weight for longer horizons. Although the optimization problem becomes more technically involved in this context, we can solve it up to the solutions to ODEs. Due to path dependence, the optimal portfolio has to be given piecewise. Proposition 6.3 summarizes the optimal portfolio weight.

Proposition 6.3. *When* $(n-1)\tau \leq T-t \leq n\tau$, $n = 1, 2, \cdots$, *the optimal wealth fraction invested in the risky asset is given by* $\phi_0^* = \phi_t^M + \phi_t^{sH} + \phi_t^{pH}$, *where*

$$\phi_t^M = \frac{\alpha m_t + (1-\alpha)\mu}{\gamma\sigma^2}, \quad \phi_t^{sH} = A_{11t}^{(n)} s_t, \quad \phi_t^{pH} = \sum_{i=2}^{2^n-1} A_{i1t}^{(n)} \tilde{B}_{it}^{(n)} + A_{1t}^{(n)}, \tag{6.7}$$

and the value function is given by

$$J(t, \tilde{B}_i^{(n)} : i = 1, \cdots, 2^n) = \frac{W_t^{1-\gamma}}{1-\gamma} \exp\left\{\gamma\left(\frac{1}{2}\sum_{i=1}^{2^n-1}\sum_{j=1}^{2^n-1} A_{ij,t}^{(n)} \tilde{B}_{i,t}^{(n)} \tilde{B}_{j,t}^{(n)} + \sum_{i=1}^{2^n-1} A_{i,t}^{(n)} \tilde{B}_{i,t}^{(n)} + \tilde{B}_{2^n,t}^{(n)} + A_{2^n,t}^{(n)}\right)\right\},$$

where the coefficients satisfy

$$\dot{A}_{ij,t}^{(n)} = \mathbb{F}_{ij}^{(n)}\big(t, A_{kl,t}^{(n)} :\ k, l = 1, \cdots, 2^n - 1\big),\ i, j = 1, \cdots, 2^n - 1,$$

$$\dot{A}_{i,t}^{(n)} = \mathbb{F}_{i}^{(n)}\big(t, s_{t-\tau}, A_{k,t}^{(n)} :\ k = 1, \cdots, 2^n - 1\big),\ i = 1, \cdots, 2^n,$$

$\mathbb{F}_{ij}^{(n)}$ *is a quadratic function of* $A_{kl,t}^{(n)}$, *and* $A_{kl,t}^{(n)} = A_{lk,t}^{(n)}$; $\mathbb{F}_i^{(n)}$ *for* $i = 1, \cdots, 2n-1$ *is a linear function of* $s_{t-\tau}$ *and* $A_{k,t}^{(n)}$; *and* $\mathbb{F}_{2n}^{(n)}$ *is a quadratic function of* $s_{t-\tau}$ *and* $A_{1,t}^{(n)}$ *and linear in* $A_{k,t}^{(n)} : k = 2, \cdots, 2^n - 1$; *and* $\tilde{B}_i^{(n)}$ *is stochastic and satisfies*

$$\tilde{B}_{1,t}^{(n)} = s_t, \quad \dot{\tilde{B}}_{i,t}^{(n)} = \mathbb{G}_i^{(n)}(t, \tilde{B}_{j,t}^{(n)} : j = 1, \cdots, 2^n - 1),\ i = 2, \cdots, 2^n.$$

$\mathbb{G}_i^{(n)}$ *for* $i = 2, \cdots, 2^n - 1$ *is linear in* $\tilde{B}_{j,t}^{(n)}$ *and* $\mathbb{G}_{2^n}^{(n)}$ *is a quadratic function of* $\tilde{B}_{j,t}^{(n)}$.

Proposition 6.3 reduces to Proposition 3.3 when $T - t \leq \tau$. In this case, s is a sufficient statistic of ϕ^*, and there is no new variable $\tilde{B}^{(1)}$ for hedging demand in (6.7).

For longer investment horizons, the optimal portfolio weight is also of the form $\phi_t^* = \phi_t^M + \phi_t^H$, where the mean-variance portfolio weight ϕ_t^M depends on momentum, and the hedging portfolio weight ϕ_t^H depends on a new horizon-dependent variable that is a weighted average of the historical instantaneous returns over the look-back period. Path dependence is also a salient feature of the optimal portfolio weight in the context of longer horizons $T - t > \tau$. Indeed, the path dependence of future returns over long horizons exhibits similar patterns as those over short horizons, as shown in Online Appendix I. (Online Appendix III.10 also provides a Monte Carlo simulation method, and we verify that the numerical solutions produced with this method are close to the closed-form solutions for the case $T - t \leq \tau$. However, the case with long investment horizon is much more difficult to study due to huge computational costs.)

6.4. **More Discussions.** A general model of momentum m_t can be given by

$$m_t = \int_{-\infty}^{t} h(t-u)\frac{dS_u}{S_u}, \tag{6.8}$$

where $h(\cdot)$ is a deterministic function on $[0,\infty)$. In (6.8), momentum is a weighted average of historical returns with the weights $h(\cdot)$. In this paper, we focus on model (2.2), which is a specification of (6.8) with $h(\cdot)$ given by $h(s) = \frac{1}{\tau}1_{s\in[0,\tau]}$. Model (4.2) is another specification of (6.8) with $h(s) = \frac{1}{\tau_i}e^{-s/\tau_i}$. Function $h(\cdot)$ plays a key role in determining the relative power of past returns in terms of predicting future returns in (2.1) and (6.8). For example, a decreasing $h(\cdot)$ implies that the recent returns of an asset have stronger predicting power than the distant returns. Different $h(\cdot)$ can lead to distinctly different return dynamics.

Most weighting schemes $h(\cdot)$ generate non-Markovian price dynamics, and only a few cases are Markovian. In a Markovian price model, the original state variables that describe the price dynamics are a sufficient statistic of the dynamics of the optimal wealth and portfolio. With these state variables, one can solve the dynamic portfolio choice problem using standard methods, e.g., the dynamic programming method. However, for a non-Markovian price model, one needs more state variables to characterize the optimal wealth and portfolio. A challenging (perhaps the most challenging) part of solving a non-Markovian stochastic control problem is constructing the sufficient statistic. Unfortunately, there seems to be no unified solution for the sufficient statistic in the case of an arbitrary function $h(\cdot)$. As a result, there is no standard way to solve for the optimal portfolio in the case of the general problem (2.1) and (6.8); some instances of this problem may not even have a solution. One can only study and compare the specific forms of the general model (6.8), as did in our paper.

7. Conclusion

In this paper, we explicitly solve for the optimal dynamic trading strategy between a riskless asset and a risky momentum asset. The optimal portfolio weight of the risky asset depends on momentum as well as historical prices of the asset; for a given level of momentum, investors may short the risky asset if it has rebound price paths but leverage if it has hump-shaped price paths. Path dependence is represented with explicit formulas as well as heuristic statistics. In contrast, the optimal portfolio weight in Merton's (1971) framework is completely determined by the momentum. The optimal strategy outperforms the mean-variance strategy, which is widely used in the literature and ignores path dependence, especially after extreme periods with large price swings.

Many important problems in economics and finance involve moving averages (MAs), such as time to build (Kydland and Prescott, 1982), post-earnings announcement drift (Bernard and Thomas, 1989), price delays (Hou and Moskowitz, 2005), and long memory in return volatility (Andersen, Bollerslev and Diebold, 2007). MAs are also widely used for financial practices, such as defining the strike price of options in the context of energy markets and corporate finance (Bouaziz, Briys and Crouhy, 1994; Dai, Li and Zhang, 2010) and computing intra-day currency returns (Melvin and Prins, 2015). Our results can also be used in these settings.

Appendix A. Proofs of Lemmas 3.1 and 3.2 and Proposition 3.3

A.1. **Proof of Lemmas 3.1 and 3.2.** It follows from (2.1) that the market price of risk is given by (3.3), which satisfies Novikov's condition $\mathbb{E}[\exp\{\frac{1}{2}\int_0^T \theta_t^2 dt\}] < \infty$. Thus, the state price density is given by (3.2). Define ξ_t as (3.4), which is a martingale under $\mathbb{P}$. The wealth process follows $dW_t = W_t(r + \sigma\theta_t\phi_t)dt + \sigma W_t\phi_t dB_t$, where ϕ_t is the portfolio weight of the risky asset. Define the martingale measure $\mathbb{Q}$ as $d\mathbb{Q}/d\mathbb{P} = \xi_T$. Under the martingale measure, the wealth process W_t follows

$$e^{-rt}W_t = W_0 + \sigma\int_0^t e^{-ru}W_u\phi_u dB_u^Q, \quad 0 \le t \le T, \tag{A.1}$$

where $B_t^Q = B_t + \int_0^t \theta_u du$ is a Brownian motion under $\mathbb{Q}$. The budget constraint can be given as $\mathbb{E}_0[\pi_T W_T] \le W_0$. Then, the problem is reduced to the unconstrained maximization of $\mathbb{E}_0[\frac{W_T^{1-\gamma}}{1-\gamma}] + \lambda(W_0 - \mathbb{E}_0[\pi_T W_T])$, where λ is the Lagrange multiplier. Proofs of this well-known result can be found in the works of Harrison and Kreps (1979), Cox and Huang (1989), and Karatzas and Shreve (1998). The first order condition leads to the following optimal terminal wealth calculation:

$$W_T = (\lambda\pi_T)^{-1/\gamma}. \tag{A.2}$$

Define $\bar{\xi}_0 = \mathbb{E}_0\big[\xi_T^{(\gamma-1)/\gamma}\big]$. Then $W_0 = \mathbb{E}_0[\pi_T W_T] = \mathbb{E}_0[\pi_T^{(\gamma-1)/\gamma}]\lambda^{-1/\gamma} = \bar{\xi}_0 e^{(1-\gamma)rT/\gamma}\lambda^{-1/\gamma}$; hence, the Lagrange multiplier is given by $\lambda = \bar{\xi}_0^\gamma W_0^{-\gamma} e^{(1-\gamma)rT}$. It follows from (A.2) that the value function satisfies

$$J = \mathbb{E}_0\left[\frac{W_T^{1-\gamma}}{1-\gamma}\right] = \frac{1}{1-\gamma}W_0^{1-\gamma}\bar{\xi}_0^\gamma e^{(1-\gamma)rT}. \tag{A.3}$$

The optimal wealth process is then given by

$$W_t = \pi_t^{-1}\mathbb{E}_t[\pi_T W_T] = W_0\bar{\xi}_0^{-1}e^{rt}\xi_t^{-1}\mathbb{E}_t\big[\xi_T^{(\gamma-1)/\gamma}\big]. \tag{A.4}$$

It follows from (A.1) that $d(e^{-rt}W_t) = \sigma e^{-rt}\phi_t W_t dB_t^Q$. In addition, Ito's formula implies that $d(e^{-rt}W_t) = d(\pi_t\xi_t^{-1}W_t) = \xi_t^{-1}(\pi_t\theta_t W_t + \psi_t)dB_t^Q$, where ψ_t is governed by $\pi_t W_t = W_0 + \int_0^t \psi_u dB_u$. By matching the volatility, the optimal portfolio weight is given by (3.6).

A.2. **Proof of Proposition 3.3.** We rewrite (2.1) as $ds_t = [(1-\alpha)(r + \mu - \frac{\sigma^2}{2}) + \frac{\alpha}{\tau}(s_t - s_{t-\tau})]dt + \sigma dB_t$, where $s_t = \ln S_t$ and hence $\theta_t = \frac{1}{\sigma}[(1-\alpha)\mu - \alpha(r - \frac{\sigma^2}{2})] + \frac{\alpha}{\tau\sigma}(s_t - s_{t-\tau})$. We define a new measure: $\frac{d\mathbb{P}^*}{d\mathbb{P}} = \exp\big\{-\int_t^T \frac{\gamma-1}{\gamma}\theta_u dB_u - \int_t^T \frac{(\gamma-1)^2}{2\gamma^2}\theta_u^2 du\big\}$; under this measure,

$$ds_t = \left[\Big(r - \frac{\sigma^2}{2}\Big)\Big(1 - \frac{\alpha}{\gamma}\Big) + (1-\alpha)\frac{\mu}{\gamma} + \frac{\alpha}{\gamma\tau}(s_t - s_{t-\tau})\right]dt + \sigma dB_t^*, \tag{A.5}$$

and

$$\mathbb{E}_0[\xi_T^{\frac{\gamma-1}{\gamma}}] = \mathbb{E}_0^*\left[\exp\left\{\frac{1-\gamma}{2\gamma^2\sigma^2}\int_0^T\left[(1-\alpha)\mu - \alpha\left(r - \frac{\sigma^2}{2}\right) + \frac{\alpha}{\tau}(s_u - s_{u-\tau})\right]^2 du\right\}\right]. \tag{A.6}$$

Suppose that $0 \leq T \leq \tau$. When $t \leq \tau$, $s_{t-\tau}$ is a realized log price and is known at time 0. Thus, s_t in (A.5) can be treated as a Markov process. Denote

$$f(s,t) = \mathbb{E}_t^*\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_t^T \frac{1}{\sigma^2}\left[(1-\alpha)\mu - \alpha\left(r - \frac{\sigma^2}{2}\right) + \frac{\alpha}{\tau}(s_u - s_{u-\tau})\right]^2 du\right\}\right]. \tag{A.7}$$

The Feynman-Kac formula implies that

$$\begin{aligned} &\frac{\partial f}{\partial t} + \left[\left(r - \frac{\sigma^2}{2}\right)\left(1 - \frac{\alpha}{\gamma}\right) + (1-\alpha)\frac{\mu}{\gamma} + \frac{\alpha}{\gamma\tau}(s - s_{t-\tau})\right]\frac{\partial f}{\partial s} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial s^2} \\ &+ \frac{1-\gamma}{2\gamma^2\sigma^2}\left[(1-\alpha)\mu - \alpha\left(r - \frac{\sigma^2}{2}\right) + \frac{\alpha}{\tau}(s - s_{t-\tau})\right]^2 f = 0. \end{aligned} \tag{A.8}$$

Its solution is given by

$$\begin{aligned} f(s_t, t) &= \exp\{A_{1,t}s_t^2/2 + A_{2,t}s_t + A_{3,t}\}, \quad \text{where} \\ \dot{A}_{1,t} &= -\sigma^2(A_{1,t})^2 - \frac{2\alpha}{\gamma\tau}A_{1,t} - \frac{1-\gamma}{\gamma^2}\frac{\alpha^2}{\tau^2\sigma^2}, \\ \dot{A}_{2,t} &= -\left(\sigma^2 A_{1,t} + \frac{\alpha}{\gamma\tau}\right)A_{2,t} - \left[(1-\alpha)\frac{\mu}{\gamma} + \left(r - \frac{\sigma^2}{2}\right)\left(1 - \frac{\alpha}{\gamma}\right)\right. \\ &\quad \left. - \frac{\alpha}{\gamma\tau}s_{t-\tau}\right]A_{1,t} - \frac{1-\gamma}{\gamma^2}\frac{\alpha}{\sigma^2\tau}\left[(1-\alpha)\mu + \alpha\left(\frac{\sigma^2}{2} - r\right) - \frac{\alpha}{\tau}s_{t-\tau}\right], \\ \dot{A}_{3,t} &= -\frac{\sigma^2}{2}A_{2,t}^2 - \frac{\sigma^2}{2}A_{1,t} - \left[(1-\alpha)\frac{\mu}{\gamma} + \left(r - \frac{\sigma^2}{2}\right)\left(1 - \frac{\alpha}{\gamma}\right) - \frac{\alpha}{\gamma\tau}s_{t-\tau}\right]A_{2,t} \\ &\quad - \frac{1-\gamma}{2\gamma^2\sigma^2}\left[(1-\alpha)\mu + \alpha\left(\frac{\sigma^2}{2} - r\right) - \frac{\alpha}{\tau}s_{t-\tau}\right]^2, \end{aligned} \tag{A.9}$$

and terminal conditions are $A_{1,T} = A_{2,T} = A_{3,T} = 0$. Here, the realized price $s_{t-\tau}$ in A_2 and A_3 is continuous but non-differentiable. Substituting (A.9) into (A.4), we have

$$\begin{aligned} dW_t/W_t &= [\theta_t^2/\gamma + \sigma\theta_t(A_{1,t}s_t + A_{2,t}) + r]dt + [\theta_t/\gamma + \sigma(A_{1,t}s_t + A_{2,t})]dB_t, \\ \psi_t &= W_t\pi_t[(1-\gamma)\theta_t/\gamma + \sigma(A_{1,t}s_t + A_{2,t})]. \end{aligned} \tag{A.10}$$

Thus, the optimal portfolio weight at time 0 is given by $\phi_t^* = \phi_t^M + \phi_t^H$, where

$$\phi_t^M = \theta_t/(\gamma\sigma) \qquad \text{and} \qquad \phi_t^H = A_{1,t}s_t + A_{2,t}, \tag{A.11}$$

and the optimal wealth process is given by

$$W_t^* = W_0\bar{\xi}_0^{-1}e^{rt}\xi_t^{-1/\gamma}\exp\{A_{1,t}(\ln S_t)^2/2 + A_{2,t}\ln S_t + A_{3,t}\}. \tag{A.12}$$

The hedging demand in (A.11) is given in terms of historical prices. Now we rewrite it in terms of historical returns. The solution to the ODEs in (A.9) is given by

$$
\begin{aligned}
A_{1,t} &= \frac{\alpha(\gamma-1)\big(1-e^{\frac{2\alpha(T-t)}{\sqrt{\gamma}\tau}}\big)}{\gamma\sigma^2\tau\Big[(\sqrt{\gamma}-1)e^{\frac{2\alpha(T-t)}{\sqrt{\gamma}\tau}}+(\sqrt{\gamma}+1)\Big]}, \\
A_{2,t} &= \int_t^T e^{\int_t^u(\sigma^2 A_{1,v}+\frac{\alpha}{\gamma\tau})dv}\Bigg[\Bigg(\Big(r-\frac{\sigma^2}{2}\Big)\Big(1-\frac{\alpha}{\gamma}\Big)+(1-\alpha)\frac{\mu}{\gamma}-\frac{\alpha}{\gamma\tau}\ln S_{u-\tau}\Bigg)A_{1,u} \\
&\quad +\frac{1-\gamma}{\gamma^2}\frac{\alpha}{\sigma^2\tau}\Bigg((1-\alpha)\mu-\alpha\Big(r-\frac{\sigma^2}{2}\Big)-\frac{\alpha}{\tau}\ln S_{u-\tau}\Bigg)\Bigg]du, \qquad \text{(A.13)}\\
A_{3,t} &= \int_t^T\Bigg[\frac{\sigma^2}{2}A_{2,u}^2+\Bigg(\Big(r-\frac{\sigma^2}{2}\Big)\Big(1-\frac{\alpha}{\gamma}\Big)+(1-\alpha)\frac{\mu}{\gamma}-\frac{\alpha}{\gamma\tau}\ln S_{u-\tau}\Bigg)A_{2,u} \\
&\quad +\frac{\sigma^2}{2}A_{1,u}+\frac{1-\gamma}{2\gamma^2\sigma^2}\Bigg((1-\alpha)\mu-\alpha\Big(r-\frac{\sigma^2}{2}\Big)-\frac{\alpha}{\tau}\ln S_{u-\tau}\Bigg)^2\Bigg]du,
\end{aligned}
$$

where $A_{2,t}$ and $A_{3,t}$ depend on price paths over $[-\tau,0]$. After some manipulations, substituting (A.13) into (A.11) produces (3.8), where

$$
\begin{aligned}
\omega_v &= \begin{cases}\int_{-\tau}^{v}\hat{\omega}_u du, & v\in[-\tau,-\tau+T],\\ \int_{-\tau}^{-\tau+T}\hat{\omega}_u du, & v\in[-\tau+T,0],\end{cases}\\
\hat{\omega}_u &= \frac{\gamma\sigma^2}{\alpha}C_{0,u+\tau}\exp\Big\{\int_0^{u+\tau}\Big[\frac{\gamma(1-\gamma)\sigma^2\tau}{\alpha}C_{0,\hat{u}}+\frac{\alpha}{\tau}\Big]d\hat{u}\Big\}>0,\\
C_{0,u} &= \frac{\alpha^2\big(e^{\frac{2\alpha(T-u)}{\sqrt{\gamma}\tau}}+1\big)}{\gamma^{3/2}\sigma^2\tau^2\Big[(\sqrt{\gamma}-1)e^{\frac{2\alpha(T-u)}{\sqrt{\gamma}\tau}}+(\sqrt{\gamma}+1)\Big]}, \qquad \text{(A.14)}\\
C_1 &= \int_0^T\exp\Big\{\int_0^u\Big[\frac{\gamma(1-\gamma)\sigma^2\tau}{\alpha}C_{0,\hat{u}}+\frac{\alpha}{\tau}\Big]d\hat{u}\Big\}\Big\{\Big[\frac{(1-\alpha)\mu-\alpha r}{\gamma}+r\\
&\quad -\frac{\sigma^2}{2}\Big]\frac{\gamma(1-\gamma)\tau}{\alpha}C_{0,u}+(r-\frac{\sigma^2}{2})\frac{(\gamma-1)\alpha}{\gamma\sigma^2\tau}\Big\}du+(1-\gamma)r\frac{\alpha}{\gamma\sigma^2}\int_{-\tau}^0\omega_v dv.
\end{aligned}
$$

Acknowledgements

We would like to thank Steven Kou (area editor), an Associate Editor, two anonymous referees, Daniel Andrei (discussant), Michael Brennan, Peter Carr, Georgy Chabakauri, Hui Chen (discussant), Li Chen, Phil Dybvig, Xue-Zhong He, Ou Hu (discussant), Ravi Jagannathan, Yuelei Li (discussant), Bernt Øksendal, Maureen O'Hara, and Guofu Zhou, as well as the conference and seminar participants at the 2016 Workshop on Complex Information and Financial Market Dynamics, the 2017 IFS Conference of Investments and Asset Pricing, the 2017 China International Conference in Finance, the 2017 Financial Management Association Annual Meeting, the 2018 Western Finance Association Meeting, Harbin Institute of Technology, Queensland University of Technology, Shanghai Advanced Institute of Finance, Southwestern University of

Finance and Economics, University of Amsterdam, University of California San Diego, and University of Technology Sydney for their helpful comments. The financial support we received from the Australian Research Council (Project ID: DE180100649) is gratefully acknowledged. A previous version of this paper was circulated under the title "Reversing Momentum: The Optimal Dynamic Momentum Strategy". All the remaining errors are ours.

References

Andersen, T., Bollerslev, T. and Diebold, F. (2007), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *Review of Economics and Statistics* **89**, 701–720.

Asness, C., Liew, J. and Stevens, R. (1997), 'Parallels between the cross-sectional predictability of stock and country returns', *Journal of Portfolio Management* **23**, 79–87.

Asness, C., Moskowitz, T. and Pedersen, L. (2013), 'Value and momentum everywhere', *Journal of Finance* **68**, 929–985.

Bernard, V. and Thomas, J. (1989), 'Post-earnings-announcement drift: Delayed price response or risk premium', *Journal of Accounting Research* **27**, 1–36.

Bouaziz, L., Briys, E. and Crouhy, M. (1994), 'The pricing of forward-starting asian options', *Journal of Banking and Finance* **18**, 823–839.

Brock, W., Lakonishok, J. and LeBaron, B. (1992), 'Simple technical trading rules and the stochastic properties of stock returns', *Journal of Finance* **47**, 1731–1764.

Burghardt, G. and Walls, B. (2011), *Managed Futures for Institutional Investors: Analysis and Portfolio Construction*, John Wiley & Sons, Hoboken, New Jersey.

Cont, R. and Fournié, D.-A. (2013), 'Functional Ito calculus and stochastic integral representation of martingales', *Annals of Probability* **41**, 109–133.

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

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

Da, Z., Gurun, U. and Warachka, M. (2014), 'Frog in the pan: Continuous information and momentum', *Review of Financial Studies* **27**, 2171–2218.

Dai, M., Li, P. and Zhang, J. (2010), 'A lattice algorithm for pricing moving average barrier options', *Journal of Economic Dynamics and Control* **34**, 542–554.

Dai, M., Yang, Z., Zhang, Q. and Zhu, Q. (2016), 'Optimal trend following trading rules', *Mathematics of Operations Research* **41**, 626–642.

Dai, M., Zhang, Q. and Zhu, Q. (2010), 'Trend following trading under a regime switching model', *SIAM Journal of Financial Mathematics* **1**, 780–810.

Daniel, K. and Moskowitz, T. (2016), 'Momentum crashes', *Journal of Financial Economics* **122**, 221–247.

Ehsani, S. and Linnainmaa, J. (2021), 'Factor momentum and the momentum factor', *Journal of Finance* **forthcoming**.

Fama, E. and French, K. (1988), 'Permanent and temporary components of stock prices', *Journal of Political Economy* **96**, 246–273.

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

Fama, E. and MacBeth, J. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**, 607–636.

George, T. and Hwang, C. (2004), 'The 52-week high and momentum investing', *Journal of Finance* **59**, 2145–2176.

Georgopoulou, A. and Wang, J. (2017), 'The trend is your friend: Time-series momentum strategies across equity and commodity markets', *Review of Finance* **21**, 1557–1592.

Gorton, G., Hayashi, F. and Rouwenhorst, K. (2013), 'The fundamentals of commodity futures returns', *Review of Finance* **17**, 35–105.

Goyal, A. and Jegadeesh, N. (2018), 'Cross-sectional and time-series tests of return predictability: What is the difference?', *Review of Financial Studies* **31**, 1784–1824.

Grinblatt, M. and Moskowitz, T. (2004), 'Predicting stock price movements from past returns: The role of consistency and tax-loss selling', *Journal of Financial Economics* **71**, 541–579.

Harrison, J. and Kreps, D. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic Theory* **20**, 381–408.

Haugh, M., Kogan, L. and Wang, J. (2006), 'Evaluating portfolio policies: A duality approach', *Operations Research* **54**, 405–418.

He, X. and Li, K. (2015), 'Profitability of time series momentum', *Journal of Banking and Finance* **53**, 140–157.

Heston, S. and Sadka, R. (2008), 'Seasonality in the cross-section of stock returns', *Journal of Financial Economics* **87**, 418–445.

Hommes, C. and Zhu, M. (2014), 'Behavioral learning equilibria', *Journal of Economic Theory* **150**, 778–814.

Hou, K. and Moskowitz, T. (2005), 'Market frictions, price delay, and the cross-section of expected returns', *Review of Financial Studies* **18**, 981–1020.

Hurst, B., Ooi, Y. and Pedersen, L. (2010), 'Understanding managed futures', *AQR Capital Management* .

Hurst, B., Ooi, Y. and Pedersen, L. (2017), 'A century of evidence on trend-following investing', *Journal of Portfolio Management* **44**, 15–29.

Jegadeesh, N. and Titman, S. (1993), 'Returns to buying winners and selling losers: Implications for stock market efficiency', *Journal of Finance* **48**, 65–91.

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

Karatzas, I. and Shreve, S. (1998), *Methods of Mathematical Finance*, Springer-Verlag, New York.

Kim, T. and Omberg, E. (1996), 'Dynamic nonmyopic portfolio behavior', *Review of Financial Studies* **9**, 141–161.

Koijen, R., Rodríguez, J. and Sbuelz, A. (2009), 'Momentum and mean reversion in strategic asset allocation', *Management Science* **55**, 1199–1213.

Kydland, F. and Prescott, E. (1982), 'Time to build and aggregate fluctuations', *Econometrica* **50**, 1345–1370.

Lewellen, J. (2015), 'The cross-section of expected stock returns', *Critical Finance Review* **4**, 1–44.

Li, J. and Yu, J. (2012), 'Investor attention, psychological anchors, and stock return predictability', *Journal of Financial Economics* **104**, 401–419.

Li, K. and Liu, J. (2018), Portfolio selection under time delays: A piecewise dynamic programming approach, working paper, Available at SSRN: https://ssrn.com/abstract=2916481.

Liu, H. and Loewenstein, M. (2002), 'Optimal portfolio selection with transaction costs and finite horizons', *Review of Financial Studies* **15**, 805–835.

Liu, J. (2007), 'Portfolio selection in stochastic environments', *Review of Financial Studies* **20**, 1–39.

Lo, A. and Hasanhodzic, J. (2010), *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*, John Wiley & Sons, Hoboken, NJ, USA.

Longstaff, F. and Schwartz, E. (2001), 'Valuing American options by simulation: A simple least-squares approach', *Review of Financial Studies* **14**, 113–147.

Makarov, I. and Rytchkov, O. (2012), 'Forecasting the forecasts of others: Implications for asset pricing', *Journal of Economic Theory* **147**, 941–966.

Melvin, M. and Prins, J. (2015), 'Equity hedging and exchange rates at the London 4 p.m. fix', *Journal of Financial Markets* **22**, 50–72.

Merton, R. (1971), 'Optimum consumption and portfolio rules in a continuous-time model', *Journal of Economic Theory* **3**, 373–413.

Moskowitz, T. and Grinblatt, M. (1999), 'Do industries explain momentum?', *Journal of Finance* **54**, 1249–1290.

Moskowitz, T., Ooi, Y. and Pedersen, L. (2012), 'Time series momentum', *Journal of Financial Economics* **104**, 228–250.

Narang, R. (2013), *Inside the Black Box: A Simple Guide to Quantitative and High Frequency Trading*, John Wiley & Sons, Hoboken, New Jersey.

Neely, C., Rapach, D., Tu, J. and Zhou, G. (2014), 'Forecasting the equity risk premium: The role of technical indicators', *Management Science* **60**, 1772–1791.

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

Okunev, J. and White, D. (2003), 'Do momentum-based strategies still work in foreign currency markets?', *Journal of Financial and Quantitative Analysis* **38**, 425–447.

Poterba, J. and Summers, L. (1988), 'Mean reversion in stock prices: Evidence and implications', *Journal of Financial Economics* **22**, 27–59.

Rachev, S., Jašić, T., Stoyanov, S. and Fabozzi, F. (2007), 'Momentum strategies based on rewardrisk stock selection criteria', *Journal of Banking and Finance* **31**, 2325–2346.

Rouwenhorst, K. (1998), 'International momentum strategies', *Journal of Finance* **53**, 267–284.

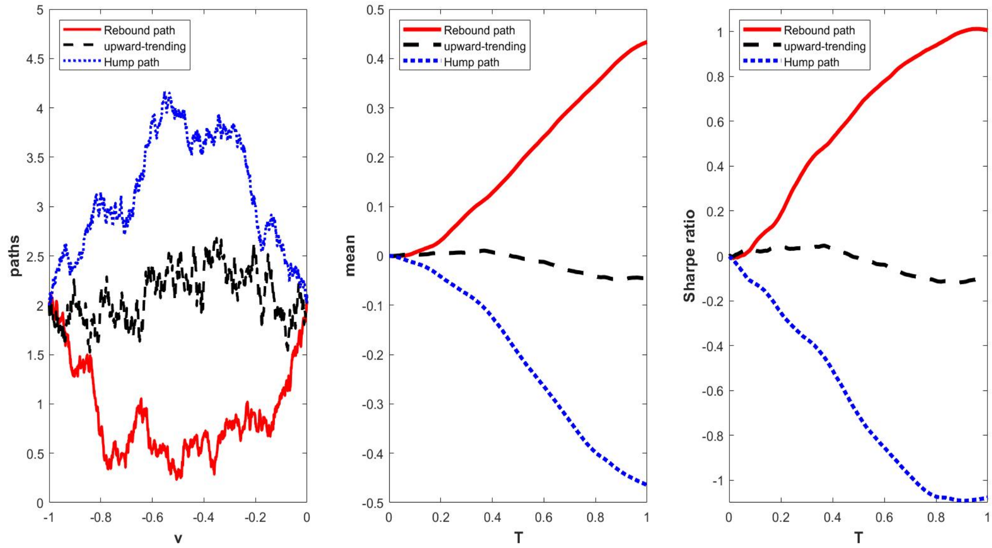

Figure 1. The left panel plots the logarithms of three typical historical price paths over the look-back period $\left(\ln S_v \text{ for } v \in [t-\tau, t]\right)$ generated by the momentum model. One has a rebound shape (red solid line), one has an upward-trending shape (black dashed line), and the other has a hump-like shape (blue dotted line). These paths have the same level of momentum ($m_0 = \mu$) and the same beginning and ending prices. The middle and right panels plot the term structures of the expected cumulative returns $\mathbb{E}_0[\ln S_T - \ln S_0]$ and the Sharpe ratios, respectively. Here, $\tau = 1$, $\alpha = 0.34$, $\mu = 0.07$, and $\sigma = 0.36$ are expressed in annual terms and they are obtained based on the estimations of the size factor detailed in Online Appendix IV.

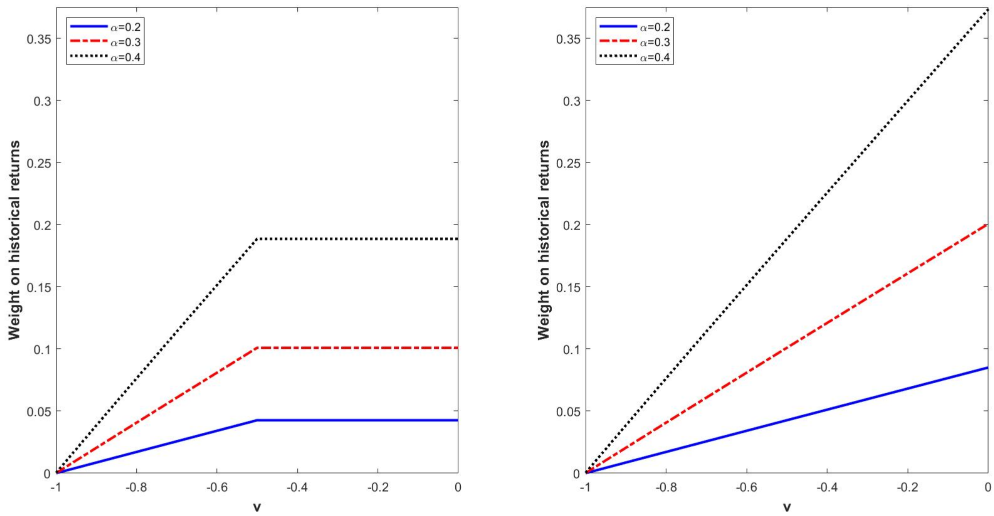

Figure 2. The weight ω_v placed by ϕ_t^H on historical instantaneous excess return $dS_v/S_v - rdv$ as a function of v for $v \in [t-\tau, t]$ for $T - t = \tau/2$ $(< \tau)$ is shown in the left panel and for $T - t = \tau$ in the right panel. Here, $\tau = 1$, $\gamma = 2$, $\mu = 0.07$, $\sigma = 0.36$, and $r = 0.04$ are in annual terms. The parameters for the momentum asset are estimated based on the size factor detailed in Online Appendix IV.

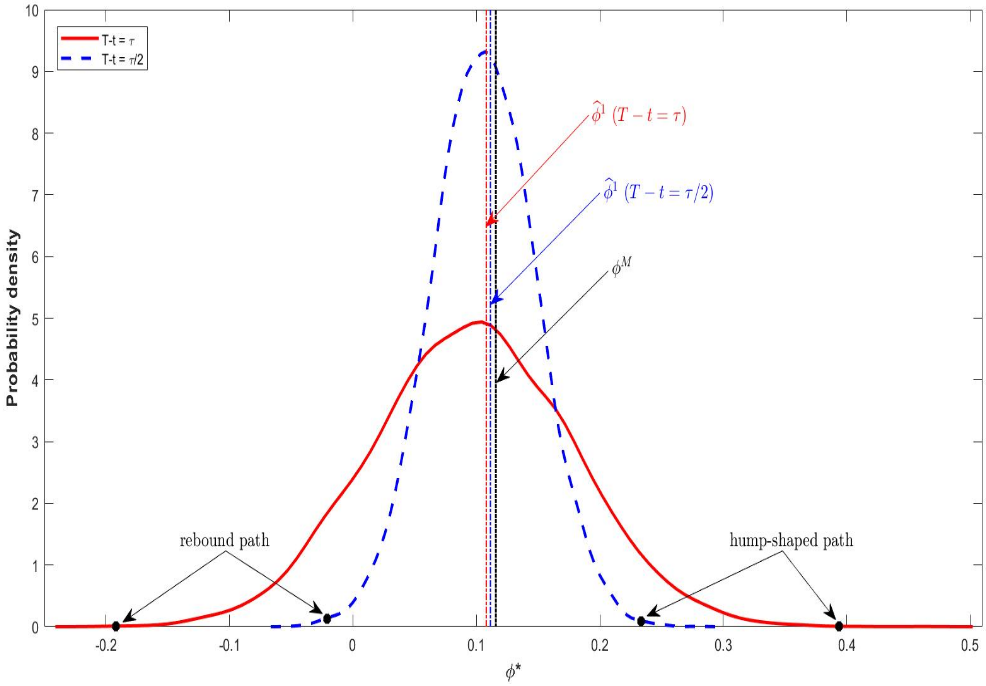

Figure 3. Given momentum ($m_t = \mu$), this figure illustrates the distributions of the optimal portfolio weights (ϕ^*) and the mean-variance portfolio weights (ϕ^M) for two different investment horizons ($T - t = \tau$ and $T - t = \tau/2$). The results are based on 10,000 simulated historical paths generated by model (2.1) and (2.2). Here, $\tau = 1$, $\gamma = 2$, $\alpha = 0.34$, $\mu = 0.07$, $\sigma = 0.36$, and $r = 0.04$ are in annual terms. The $\widehat{\phi}^1$ is the optimal portfolio weight for the historical average (HA) approximation (4.2) with $N = 1$. The parameters for the momentum asset are estimated based on the size factor detailed in Online Appendix IV.

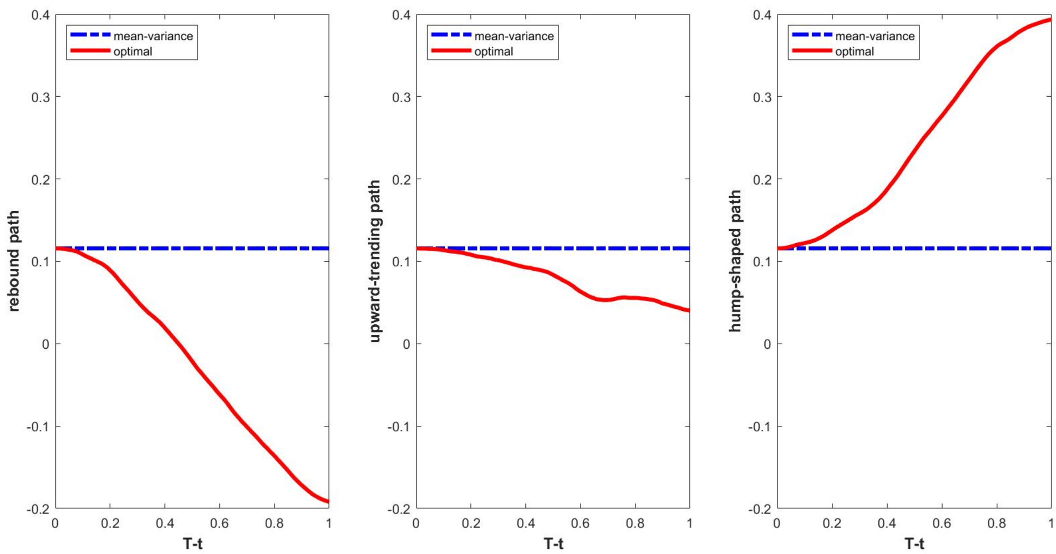

Figure 4. This figure plots the optimal portfolio weights and the mean-variance portfolio weights against the investment horizons for the three paths illustrated in Figure 1. The left panel shows the rebound path, the middle panel shows the upward-trending path, and the right panel shows the hump-shaped path. Here, $\tau = 1$, $\gamma = 2$, $\alpha = 0.34$, $\mu = 0.07$, $\sigma = 0.36$, and $r = 0.04$ are in annual terms. The parameters for the momentum asset are estimated based on the size factor detailed in Online Appendix IV.

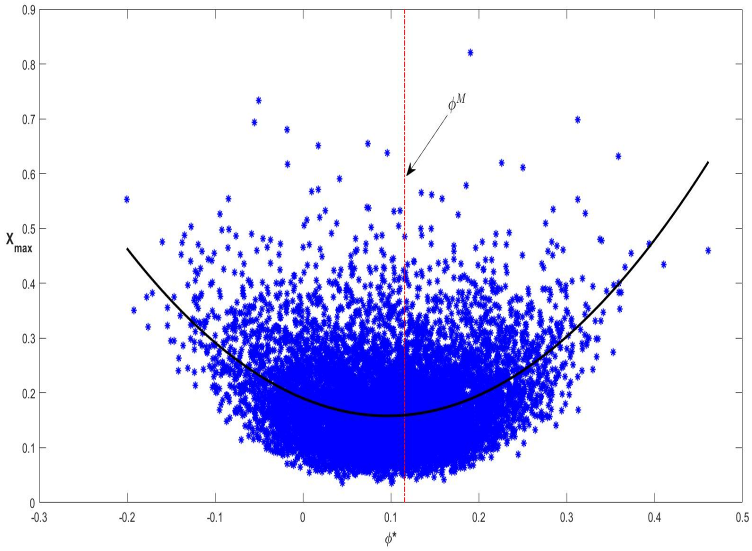

Figure 5. This figure plots the optimal portfolio weight ϕ^* against X_{max}. The solid black line illustrates the regression $X_{max} = 0.19 - 0.67 \times \phi^* + 3.48 \times (\phi^*)^2 + \epsilon$. The results are based on 10,000 simulated historical paths generated by model (2.1) and (2.2). Here, $\tau = 1$, $\gamma = 2$, $\alpha = 0.34$, $\mu = 0.07$, $\sigma = 0.36$, $r = 0.04$, and $T = 1$ are in annual terms. The parameters for the momentum asset are estimated based on the size factor detailed in Online Appendix IV.

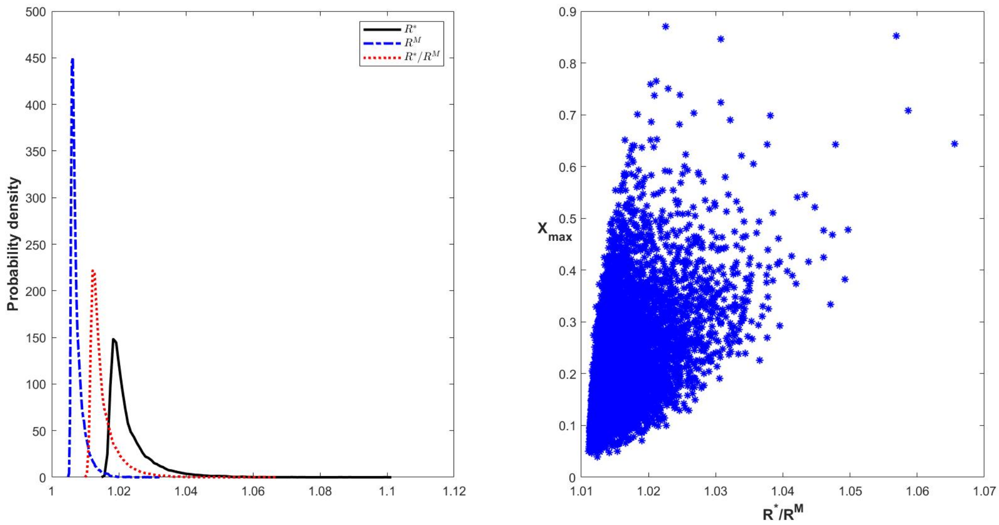

Figure 6. The left panel plots the distributions of the present value of the certainty equivalent wealth of the optimal momentum strategy R^* and the mean-variance strategy R^M, as well as the distribution of their ratios R^*/R^M. The right panel plots these ratios against X_{max}, which measures large price swings. The results are based on 10,000 simulated historical paths generated by model (2.1) and (2.2). Here $\tau = 1$, $\gamma = 2$, $\alpha = 0.34$, $\mu = 0.07$, $\sigma = 0.36$, $r = 0.04$, and $T = 1$ in annual terms. The parameters for the momentum asset are estimated based on the size factor detailed in Online Appendix IV.

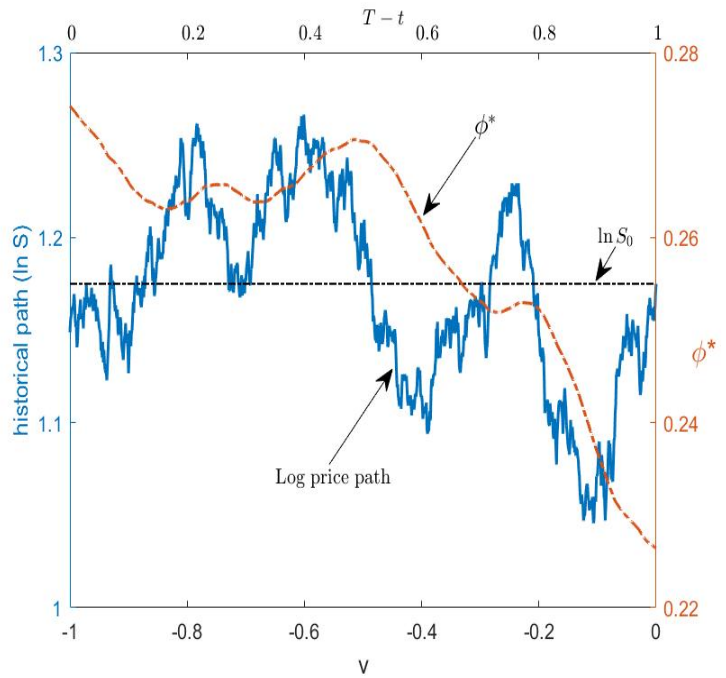

Figure 7. This figure plots the horizon dependence of the optimal portfolio weight. The solid blue line is the logarithm of the price path over the look-back period $v \in [t-\tau, t]$ generated by the momentum model. The dashed red line illustrates the optimal portfolio weights for different investment horizons $T \in [0, \tau]$. Here, $\tau = 1$, $\gamma = 2$, $\mu = 0.07$, $\alpha = 0.34$, $\sigma = 0.36$, and $r = 0.04$ are in annual terms. The parameters for the momentum asset are estimated based on the size factor detailed in Online Appendix IV.

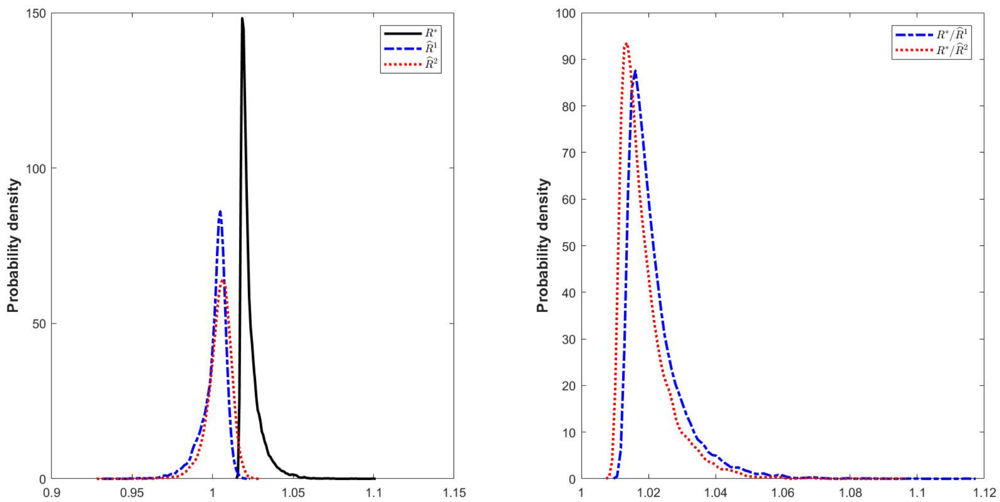

Figure 8. The left panel plots the distributions of the present value of the certainty equivalent wealth of the optimal momentum strategy R^*, the HA strategy $\hat{R}^1$ with $N = 1$, and the HA strategy $\hat{R}^2$ with $N = 2$, and the right panel plots the distribution of their ratios $R^*/\hat{R}^N$ for $N = 1, 2$. The results are based on 10,000 simulated historical paths generated by model (2.1) and (2.2). Here the parameters for the momentum model are given by $\tau = 1$, $\alpha = 0.34$, $\mu = 0.07$, and $\sigma = 0.36$, the parameters for the HA model with $N = 1$ are given by $\tau_1 = 1$, $\alpha_1 = 0.17$, $\mu = 0.08$, and $\sigma = 0.36$, and the parameters for the HA model with $N = 2$ are given by $\tau_1 = 1$, $\tau_2 = 10$, $\alpha_1 = 0.03$, $\alpha_2 = 0.36$, $\mu = 0.08$, and $\sigma = 0.36$. These parameters are estimated based on the size factor detailed in Online Appendix IV. We also set $r = 0.04$, $\gamma = 2$ and $T = 1$. All parameters are in annual terms.

ELECTRONIC COMPANIONS

for "Optimal Dynamic Momentum Strategies"

Appendix I. Return Characteristics of Momentum Assets

In this section, we examine the return characteristics of the momentum asset according to (2.1) and (2.2). Define $s_t = \ln S_t$. Due to past dependence, the expected returns, return volatility, and Sharpe ratios are given piecewise in the following proposition.

Proposition I.1. *For $T - t \in [n\tau, (n+1)\tau]$, $n = 0, 1, 2, \cdots$, the cumulative return of the asset over $[t, T]$ is given by*

$$\begin{aligned} s_T - s_t =& \frac{\tau}{\alpha}(1-\alpha)\left(r + \mu - \frac{\sigma^2}{2}\right)\left[\sum_{i=0}^{n}\left(\sum_{j=0}^{i}\frac{(-\frac{\alpha}{\tau})^j (T-t-i\tau)^j}{j!}\right)e^{\frac{\alpha}{\tau}(T-t-i\tau)} - n - 1\right] \\ & + \left[\sum_{i=0}^{n}\frac{(-\frac{\alpha}{\tau})^i (T-t-i\tau)^i}{i!}e^{\frac{\alpha}{\tau}(T-t-i\tau)} - 1\right]s_t \\ & - \frac{\alpha}{\tau}\int_{-\tau}^{0}\left[\sum_{i=1}^{n}\frac{(-\frac{\alpha}{\tau})^{i-1}(T-t-i\tau-u)^{i-1}}{(i-1)!}e^{\frac{\alpha}{\tau}(T-t-i\tau-u)}\right]s_{t+u}du \\ & - \frac{\alpha}{\tau}\int_{-\tau}^{T-t-(n+1)\tau}\left[\frac{(-\frac{\alpha}{\tau})^n [T-t-(n+1)\tau-u]^n}{n!}e^{\frac{\alpha}{\tau}[T-t-(n+1)\tau-u]}\right]s_{t+u}du \\ & + \sigma\sum_{i=0}^{n}\int_{0}^{T-t-i\tau}\frac{(-\frac{\alpha}{\tau})^i (T-t-i\tau-u)^i}{i!}e^{\frac{\alpha}{\tau}(T-t-i\tau-u)}dB_{t+u}. \end{aligned} \tag{I.1}$$

Proof. Let $s_t = \ln S_t$. Then, we have

$$m_t = \frac{1}{\tau}(s_t - s_{t-\tau}) - \left(r - \frac{\sigma^2}{2}\right), \tag{I.2}$$

and hence

$$ds_t = \left[(1-\alpha)\left(r + \mu - \frac{\sigma^2}{2}\right) + \frac{\alpha}{\tau}(s_t - s_{t-\tau})\right]dt + \sigma dB_t, \tag{I.3}$$

which implies that

$$s_t = \frac{\tau}{\alpha}(1-\alpha)\left(r + \mu - \frac{\sigma^2}{2}\right)\left[e^{\frac{\alpha}{\tau}t} - 1\right] + e^{\frac{\alpha}{\tau}t}s_0 - \frac{\alpha}{\tau}\int_{-\tau}^{t-\tau} e^{\frac{\alpha}{\tau}(t-\tau-v)}s_v dv + \sigma\int_0^t e^{\frac{\alpha}{\tau}(t-v)}dB_v. \tag{I.4}$$

We want to separate s_t into two parts: one is determined by the initial values and another collects all the innovations. Notice that the third term in (I.4) comprises the price s over

$[-\tau, t-\tau]$. For $t \in [0,\tau]$, the price is completely determined by the initial values; hence, we have

$$s_t = \frac{\tau}{\alpha}(1-\alpha)\left(r+\mu-\frac{\sigma^2}{2}\right)\left[e^{\frac{\alpha}{\tau}t}-1\right]+e^{\frac{\alpha}{\tau}t}s_0-\frac{\alpha}{\tau}\int_{-\tau}^{t-\tau}e^{\frac{\alpha}{\tau}(t-\tau-v)}s_v dv+\sigma\int_0^t e^{\frac{\alpha}{\tau}(t-v)}dB_v. \quad \text{(I.5)}$$

For $t \in [\tau, 2\tau]$, (I.4) becomes

$$\begin{aligned} s_t = &\frac{\tau}{\alpha}(1-\alpha)\left(r+\mu-\frac{\sigma^2}{2}\right)\left[e^{\frac{\alpha}{\tau}t}-1\right]+e^{\frac{\alpha}{\tau}t}s_0 \\ &-\frac{\alpha}{\tau}\int_0^{t-\tau}e^{\frac{\alpha}{\tau}(t-\tau-v)}s_v dv-\frac{\alpha}{\tau}\int_{-\tau}^{0}e^{\frac{\alpha}{\tau}(t-\tau-v)}s_v dv+\sigma\int_0^t e^{\frac{\alpha}{\tau}(t-v)}dB_v. \end{aligned} \quad \text{(I.6)}$$

Notice that for $v \in [0, t-\tau] \subseteq [0,\tau]$, s_v is given by (I.5). By replacing s_v in the third term of (I.6) with (I.5), we have

$$\begin{aligned} s_t = &\frac{\tau}{\alpha}(1-\alpha)\left(r+\mu-\frac{\sigma^2}{2}\right)\left(e^{\frac{\alpha}{\tau}t}+[1-\frac{\alpha}{\tau}(t-\tau)]e^{\frac{\alpha}{\tau}(t-\tau)}-2\right)+\left[e^{\frac{\alpha}{\tau}t}-\frac{\alpha}{\tau}(t-\tau)e^{\frac{\alpha}{\tau}(t-\tau)}\right]s_0 \\ &-\frac{\alpha}{\tau}\int_{-\tau}^{0}e^{\frac{\alpha}{\tau}(t-\tau-v)}s_v dv+\frac{\alpha^2}{\tau^2}\int_{-\tau}^{t-2\tau}(t-2\tau-v)e^{\frac{\alpha}{\tau}(t-2\tau-v)}s_v dv \\ &+\sigma\int_0^t e^{\frac{\alpha}{\tau}(t-v)}dB_v-\frac{\sigma\alpha}{\tau}\int_0^{t-\tau}(t-\tau-v)e^{\frac{\alpha}{\tau}(t-\tau-v)}dB_v. \end{aligned}$$

We can rewrite (I.4) for $t \in [n\tau, (n+1)\tau]$ as

$$\begin{aligned} s_t = &\frac{\tau}{\alpha}(1-\alpha)\left(r+\mu-\frac{\sigma^2}{2}\right)\left[e^{\frac{\alpha}{\tau}t}-1\right]+e^{\frac{\alpha}{\tau}t}s_0+\sigma\int_0^t e^{\frac{\alpha}{\tau}(t-v)}dB_v \\ &-\frac{\alpha}{\tau}\left(\int_{-\tau}^{0}+\int_0^{\tau}+\cdots+\int_{(n-1)\tau}^{t-\tau}\right)e^{\frac{\alpha}{\tau}(t-\tau-v)}s_v dv. \end{aligned} \quad \text{(I.7)}$$

By substituting s_v, $v \in [i\tau, (i+1)\tau]$, $i = 0, 1, \cdots, n-1$ into the last term of (I.7), we can separate s_t for $t \in [n\tau, (n+1)\tau]$ into a component of initial values and a component of Brownian motions. Therefore, mathematical induction implies that

$$\begin{aligned} s_t = &\frac{\tau}{\alpha}(1-\alpha)\left(r+\mu-\frac{\sigma^2}{2}\right)\left[\sum_{i=0}^{n}\left(\sum_{j=0}^{i}\frac{(-\frac{\alpha}{\tau})^j(t-i\tau)^j}{j!}\right)e^{\frac{\alpha}{\tau}(t-i\tau)}-n-1\right] \\ &+\sum_{i=0}^{n}\frac{(-\frac{\alpha}{\tau})^i(t-i\tau)^i}{i!}e^{\frac{\alpha}{\tau}(t-i\tau)}s_0-\frac{\alpha}{\tau}\int_{-\tau}^{0}\left[\sum_{i=1}^{n}\frac{(-\frac{\alpha}{\tau})^{i-1}(t-i\tau-v)^{i-1}}{(i-1)!}e^{\frac{\alpha}{\tau}(t-i\tau-v)}\right]s_v dv \\ &-\frac{\alpha}{\tau}\int_{-\tau}^{t-(n+1)\tau}\left[\frac{(-\frac{\alpha}{\tau})^n[t-(n+1)\tau-v]^n}{n!}e^{\frac{\alpha}{\tau}[t-(n+1)\tau-v]}\right]s_v dv \\ &+\sigma\sum_{i=0}^{n}\int_0^{t-i\tau}\frac{(-\frac{\alpha}{\tau})^i(t-i\tau-v)^i}{i!}e^{\frac{\alpha}{\tau}(t-i\tau-v)}dB_v, \qquad t \in [n\tau, (n+1)\tau]. \end{aligned}$$

The mean of $\ln(S_t/S_0) = s_t - s_0$ is the first four terms minus s_0. The variance is

$$\mathbb{V}\mathrm{ar}_0\big[s_t - s_0\big] = \sigma^2 \Bigg[\int_{-\tau}^{0} e^{-\frac{2\alpha}{\tau}u} du + \int_{-2\tau}^{-\tau} \left(\sum_{i=0}^{1} \frac{(-\frac{\alpha}{\tau})^i (-i\tau - u)^i}{i!} e^{\frac{\alpha}{\tau}(-i\tau - u)} \right)^2 du + \cdots$$
$$+ \int_{-n\tau}^{-(n-1)\tau} \left(\sum_{i=0}^{n-1} \frac{(-\frac{\alpha}{\tau})^i (-i\tau - u)^i}{i!} e^{\frac{\alpha}{\tau}(-i\tau - u)} \right)^2 du + \int_{-t}^{-n\tau} \left(\sum_{i=0}^{n} \frac{(-\frac{\alpha}{\tau})^i (-i\tau - u)^i}{i!} e^{\frac{\alpha}{\tau}(-i\tau - u)} \right)^2 du \Bigg]. \tag{I.8}$$

□

Proposition I.1 shows that the asset returns (I.1) over $[t, T]$ are the weighted sum of the historical prices s_u for $u \in [t - \tau, t]$. Therefore, the return process depends on the historical instantaneous returns, rather than just on the beginning and ending prices of the look-back period.

More importantly, the weights on different historical prices in (I.1) are different. This implies that different historical price paths, even if they have the same level of momentum, lead to different expected returns and different Sharpe ratios. However, volatility is independent of price paths, as shown by (I.8).

In general, the distribution of a finite-horizon return is determined by the instantaneous return because finite-horizon returns are cumulated by instantaneous returns. If an instantaneous return is Markovian, its conditional mean and variance constitute a sufficient statistic of the distribution of the related finite-horizon returns. Otherwise, the sufficient statistic involves more state variables due to path dependence. When returns exhibit momentum, this momentum is a sufficient statistic of the instantaneous returns but it is insufficient to characterize the finite-horizon returns. Thus, instantaneous and finite-horizon returns are characterized by different sets of state variables in the momentum model but the same set of state variables in Merton's (Markovian) model.

We numerically examine the path dependence of future returns in Equation (I.1). Figure I.1 shows that the future return $\ln S_T - \ln S_{T-\tau}$ places more weight on the recent historical (instantaneous) returns over the look-back period $[-\tau, 0]$ than on the older returns. More importantly, distribution of the weights of the historical returns becomes flat as T increases. This implies that although future returns depend more heavily on recent than on distant returns as the horizon T increases, the path dependence generated by future "momentum cycles" becomes increasingly weaker. Figure I.1 shows that the path dependence of a cumulative return $\ln S_T - \ln S_0$ is mainly driven by the future return $\ln S_\tau - \ln S_0$ over the first momentum cycle.

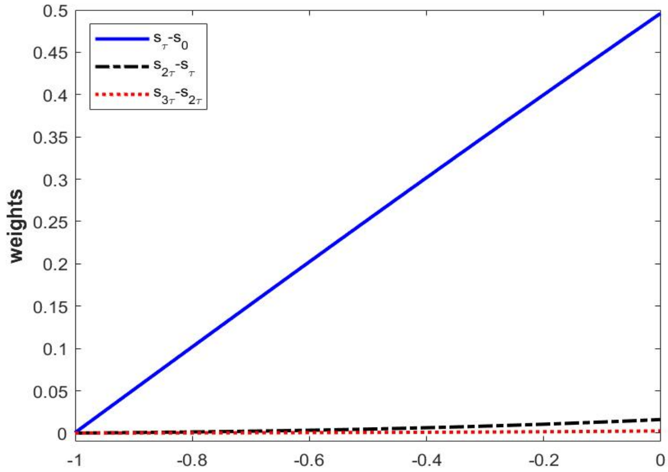

FIGURE I.1. The weights placed by $\ln S_T - \ln S_{T-\tau}$ for $T = \tau$, 2τ, and 3τ on historical instantaneous returns over the look-back period. Here $\tau = 1$ and $\alpha = 0.34$.

In addition, the weights of all the historical prices in (I.1) sum to zero. Thus, the price level of the momentum asset does not affect its returns.

Finally, momentum increases return volatility; however, its impact on expected returns depends on the related historical path, which has infinite dimensions. Although the means and variances of the returns and the Sharpe ratios are given piecewise in Proposition I.1, we find that they are continuous in our untabulated numerical simulations.

APPENDIX II. CALCULATION DETAILS FOR THE HA MODEL (4.2)

II.1. **Proof of Proposition 4.1.** For the optimization problem (3.1), the investor's value function, namely, $J(t, W, m_i)$, is governed by

$$\max_{\hat{\phi}} \Big\{ \frac{\partial J}{\partial t} + W\Big\{\hat{\phi}\Big[\sum_{i=1}^{N} \alpha_i m_i + \Big(1 - \sum_{i=1}^{N} \alpha_i\Big)\mu\Big] + r\Big\}\frac{\partial J}{\partial W} + \frac{W^2 \hat{\phi}^2 \sigma^2}{2}\frac{\partial^2 J}{\partial W^2} + \sum_{j=1}^{N} W\hat{\phi}\frac{\sigma^2}{\tau_j}\frac{\partial^2 J}{\partial W \partial m_j}$$
$$+ \sum_{j=1}^{N} \frac{1}{\tau_j}\Big[\Big(1 - \sum_{i=1}^{N}\alpha_i\Big)\mu - (m_j - \sum_{i=1}^{N}\alpha_i m_i)\Big]\frac{\partial J}{\partial m_j} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\sigma^2}{\tau_i \tau_j}\frac{\partial^2 J}{\partial m_i \partial m_j}\Big\} = 0,$$

with the boundary condition $J(T, W, m_i) = \frac{W^{1-\gamma}}{1-\gamma}$, where the wealth process satisfies $dW_t/W_t = \{\hat{\phi}_t[\sum_{i=1}^N \alpha_i m_{it} + (1 - \sum_{i=1}^N \alpha_i)\mu] + r\}dt + \sigma\hat{\phi}_t dB_t$ and $\hat{\phi}_t$ is the portfolio weight of the risky asset. The solution is given by $J = \frac{W^{1-\gamma}}{1-\gamma} f^{\gamma}$, where $f(m_i, t) = \exp\{\frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N A_{ijt} m_{it} m_{jt} + \sum_{i=1}^N A_{it} m_{it} + A_{N+1t}\}$, A_{ijt} and A_{it} are deterministic coefficients governed by the following

ODEs:

$$\dot{A}_{ijt} + \frac{2\alpha_i}{\gamma}\sum_{k=1}^{N}\frac{A_{jkt}}{\tau_k} - \frac{2}{\tau_j}A_{ijt} + \frac{1-\gamma}{\gamma^2\sigma^2}\alpha_i\alpha_j = 0, \quad i,j=1,\cdots,N,$$

$$\dot{A}_{it} + \frac{\alpha_i}{\gamma}\sum_{j=1}^{N}\frac{A_{jt}}{\tau_j} + \frac{\mu}{\gamma}\Big(1-\sum_{j=1}^{N}\alpha_j\Big)\sum_{k=1}^{N}\frac{A_{ikt}}{\tau_k} - \frac{A_{it}}{\tau_i} + \frac{1-\gamma}{\gamma^2}\frac{\mu}{\sigma^2}\Big(1-\sum_{j=1}^{N}\alpha_j\Big)\alpha_i = 0, \quad i=1,\cdots,N,$$

$$\dot{A}_{N+1t} + \frac{\mu}{\gamma}\Big(1-\sum_{i=1}^{N}\alpha_i\Big)\sum_{j=1}^{N}\frac{A_{jt}}{\tau_j} + \frac{\sigma^2}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{A_{ijt}}{\tau_i\tau_j} + \frac{1-\gamma}{\gamma}r + \frac{1-\gamma}{\gamma^2}\frac{\mu^2}{2\sigma^2}\Big(1-\sum_{i=1}^{N}\alpha_i\Big)^2 = 0, \tag{II.1}$$

with the terminal conditions $A_{ijT} = 0$ for $i,j = 1,\cdots,N$ and $A_{iT} = 0$ for $i = 1,\cdots,N+1$. The FOC leads to the optimal portfolio weight of (4.3).

II.2. **The CEW of the HA Strategy.** The optimal portfolio weight of a risky asset of which the price follows the HA approximation model (4.2) is given by

$$\widehat{\phi}_t^N = \frac{1}{\gamma\sigma^2}\Big[\sum_{i=1}^{N}\alpha_i m_{it} + \Big(1-\sum_{i=1}^{N}\alpha_i\Big)\mu\Big] + \sum_{i=1}^{N}\frac{1}{\tau_i}\Big(\sum_{j=1}^{N}A_{ijt}m_{jt} + A_{it}\Big), \tag{II.2}$$

where N is the number of Markovian state variables and A_{ijt} and A_{it} are deterministic functions given by (II.1).

When an investor applies the HA strategy (II.2) to the momentum asset in (2.1)–(2.2), her wealth process follows

$$\frac{d\hat{W}_t}{\hat{W}_t} = \{\widehat{\phi}_t[\alpha m_t + (1-\alpha)\mu] + r\}dt + \sigma\widehat{\phi}_t dB_t, \tag{II.3}$$

where $m_t = \frac{1}{\tau}(s_t - s_{t-\tau}) - r + \sigma^2/2$. We define a new measure, namely,

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \exp\Big\{(1-\gamma)\sigma\int\widehat{\phi}_t dB_t - \frac{(1-\gamma)^2\sigma^2}{2}\int\widehat{\phi}_t^2 dt\Big\}. \tag{II.4}$$

Under this new measure,

$$ds_t = \Big[\frac{\alpha}{\tau}(s_t - s_{t-\tau}) + (1-\gamma)\sigma^2\widehat{\phi}_t^N + (1-\alpha)(r + \mu - \frac{\sigma^2}{2})\Big]dt + \sigma d\tilde{B}_t,$$

$$dm_{it} = \frac{1}{\tau_i}\Big[(1-\sum_{j=1}^{N}\alpha_j)\mu - (m_{it} - \sum_{j=1}^{N}\alpha_j m_{jt}) + (1-\gamma)\sigma^2\widehat{\phi}_t^N\Big]dt + \frac{\sigma}{\tau_i}d\tilde{B}_t, \quad \text{for} \quad i=1,\cdots,N.$$

The CEW in this context satisfies

$$\begin{aligned}&\mathrm{CEW}_{\hat{\phi}^N}^{1-\gamma} = \mathbb{E}_0[\hat{W}_T^{1-\gamma}]\\ =&\tilde{\mathbb{E}}_0\Big[\exp\Big\{(1-\gamma)\int_0^T\widehat{\phi}_t^N\Big[\frac{\alpha}{\tau}(s_t - s_{t-\tau}) - \alpha r + \frac{\alpha\sigma^2}{2} + (1-\alpha)\mu\Big] + r - \frac{\gamma\sigma^2}{2}(\widehat{\phi}_t^N)^2 dt\Big\}\Big].\end{aligned}$$

Denote

$$\tilde{f}_t = \tilde{\mathbb{E}}_t\Big[\exp\Big\{(1-\gamma)\int_t^T \widehat{\phi}_u^N\Big[\frac{\alpha}{\tau}(s_u - s_{u-\tau}) - \alpha r + \frac{\alpha\sigma^2}{2} + (1-\alpha)\mu\Big] + r - \frac{\gamma\sigma^2}{2}(\widehat{\phi}_u^N)^2 du\Big\}\Big].$$

By Feynman-Kac formula, $\tilde{f}_t$ satisfies the following partial differential equation:

$$\begin{aligned}
&\frac{\partial \tilde{f}}{\partial t} + \frac{\partial \tilde{f}}{\partial s}\Big[\frac{\alpha}{\tau}(s - s_{t-\tau}) + (1-\gamma)\sigma^2\widehat{\phi}^N + (1-\alpha)(r + \mu - \frac{\sigma^2}{2})\Big] \\
&+\sum_{i=1}^N \frac{\partial \tilde{f}}{\partial m_i}\frac{1}{\tau_i}\Big[(1 - \sum_{j=1}^N \alpha_j)\mu - (m_i - \sum_{j=1}^N \alpha_j m_j) + (1-\gamma)\sigma^2\widehat{\phi}^N\Big] \\
&+\frac{\sigma^2}{2}\frac{\partial^2 \tilde{f}}{\partial s^2} + \sum_{i=1}^N \frac{\partial^2 \tilde{f}}{\partial s \partial m_i}\frac{\sigma^2}{\tau_i} + \sum_{i=1}^N \frac{\partial^2 \tilde{f}}{\partial m_i \partial m_j}\frac{\sigma^2}{2\tau_i\tau_j} \\
&+(1-\gamma)\Big\{\widehat{\phi}^N\Big[\frac{\alpha}{\tau}(s - s_{t-\tau}) - \alpha r + \frac{\alpha\sigma^2}{2} + (1-\alpha)\mu\Big] + r - \frac{\gamma\sigma^2}{2}(\widehat{\phi}^N)^2\Big\}\tilde{f} = 0,
\end{aligned} \tag{II.5}$$

with $\tilde{f}_T = 1$, where $\widehat{\phi}^N$ given by (II.2) depends on m_i $(i = 1, \cdots, N)$. Its solution is of the form:

$$\tilde{f}_t = \exp\Big\{\frac{F_{00t}}{2}s_t^2 + \sum_{i=1}^N F_{0it}s_t m_{it} + \sum_{i=1}^N\sum_{j=1}^N \frac{F_{ijt}}{2}m_{it}m_{jt} + F_{0t}s_t + \sum_{i=1}^N F_{it}m_{it} + F_{N+1t}\Big\},$$

with $F_{ij} = F_{ji}$, where F_{ij} and F_i are deterministic functions of t determined by matching the coefficients of state variables s and m_i.

In our paper, we focus on the cases of $N = 1, 2$. When $N = 1$, we obtain

$$\begin{aligned}
&\dot{F}_{00} + \frac{2\alpha}{\tau}F_{00} + \sigma^2 F_{00}^2 + \frac{2\sigma^2}{\tau_1}F_{00}F_{01} + \frac{\sigma^2}{\tau_1^2}F_{01}^2 = 0, \\
&\dot{F}_{01} + \sigma^2\hat{F}_1 F_{00} + \Big(\frac{\alpha}{\tau} + \hat{F}_5\Big)F_{01} + \sigma^2 F_{00}F_{01} + \frac{\sigma^2}{\tau_1}(F_{00}F_{11} + F_{01}^2) + \frac{\sigma^2}{\tau_1^2}F_{01}F_{11} + \frac{\alpha}{\tau}\hat{F}_1 = 0, \\
&\dot{F}_{11} + 2\sigma^2\hat{F}_1 F_{01} + 2\hat{F}_5 F_{11} + \sigma^2 F_{01}^2 + \frac{2\sigma^2}{\tau_1}F_{01}F_{11} + \frac{\sigma^2}{\tau_1^2}F_{11}^2 - \frac{\gamma\sigma^2}{1-\gamma}\hat{F}_1^2 = 0, \\
&\dot{F}_0 + \hat{F}_4 F_{00} + \hat{F}_6 F_{01} + \frac{\alpha}{\tau}F_0 + \sigma^2 F_{00}F_0 + \frac{\sigma^2}{\tau_1}(F_{00}F_1 + F_{01}F_0) + \frac{\sigma^2}{\tau_1^2}F_{01}F_1 + \frac{\alpha}{\tau}\hat{F}_3 = 0, \\
&\dot{F}_1 + \hat{F}_4 F_{01} + \hat{F}_6 F_{11} + \sigma^2\hat{F}_1 F_0 + \hat{F}_5 F_1 + \sigma^2 F_{01}F_0 + \frac{\sigma^2}{\tau_1}(F_{01}F_1 + F_{11}F_0) + \frac{\sigma^2}{\tau_1^2}F_{11}F_1 \\
&\quad + \hat{F}_1\hat{F}_2 - \frac{\gamma\sigma^2}{1-\gamma}\hat{F}_1\hat{F}_3 = 0, \\
&\quad \dot{F}_2 + \hat{F}_4 F_0 + \hat{F}_6 F_1 + \frac{\sigma^2}{2}(F_{00} + F_0^2) + \frac{\sigma^2}{\tau_1}(F_{01} + F_0 F_1) + \frac{\sigma^2}{2\tau_1^2}(F_{11} + F_1^2) + \hat{F}_2\hat{F}_3 \\
&\qquad - \frac{\gamma}{1-\gamma}\frac{\sigma^2}{2}\hat{F}_3^2 + (1-\gamma)r = 0,
\end{aligned}$$

with the terminal conditions $F_{ij,T} = F_{i,T} = 0$; additionally, the corresponding coefficients are given by

$$\begin{aligned}
\hat{F}_{1t} &= (1-\gamma)\Big(\frac{\alpha}{\gamma\sigma^2} + \frac{A_{11t}}{\tau_1}\Big),\\
\hat{F}_{2t} &= -\frac{\alpha}{\tau}s_{t-\tau} - \alpha r + \frac{\alpha\sigma^2}{2} + (1-\alpha)\mu,\\
\hat{F}_{3t} &= (1-\gamma)\Big(\frac{1-\alpha}{\gamma\sigma^2}\mu + \frac{A_{1t}}{\tau_1}\Big),\\
\hat{F}_{4t} &= -\frac{\alpha}{\tau}s_{t-\tau} + \sigma^2\hat{F}_{3t} + (1-\alpha)\Big(r + \mu - \frac{\sigma^2}{2}\Big),\\
\hat{F}_{5t} &= \frac{1-\gamma}{\gamma\tau_1}\alpha - \frac{1-\alpha}{\tau_1} + \frac{A_{11t}}{\tau_1},\\
\hat{F}_{6t} &= \frac{\sigma^2}{\tau_1}\hat{F}_{3t} + \frac{(1-\alpha)\mu}{\tau_1},
\end{aligned}$$

where A_{11t} and A_{1t} are deterministic functions given by (II.1) with $N = 1$.

When $N = 2$, we obtain

$$\dot{F}_{00} + \frac{2\alpha}{\tau}F_{00} + \sigma^2F_{00}^2 + \frac{2\sigma^2}{\tau_1}F_{00}F_{01} + \frac{2\sigma^2}{\tau_2}F_{00}F_{02} + \frac{\sigma^2}{\tau_1^2}F_{01}^2 + \frac{\sigma^2}{\tau_2^2}F_{02}^2 + \frac{2\sigma^2}{\tau_1\tau_1}F_{01}F_{02} = 0,$$

$$\begin{aligned}
&\dot{F}_{01} + (1-\gamma)\frac{\alpha}{\tau}\tilde{F}_1 + (1-\gamma)\sigma^2\tilde{F}_1F_{00} + \frac{\alpha}{\tau}F_{01} + \frac{\tilde{F}_3}{\tau_1}F_{01} + \frac{\tilde{F}_4}{\tau_2}F_{02} + \sigma^2F_{00}F_{01} + \frac{\sigma^2}{\tau_1}(F_{00}F_{11} + F_{01}^2)\\
&\quad + \frac{\sigma^2}{\tau_2}(F_{00}F_{12} + F_{01}F_{02}) + \frac{\sigma^2}{\tau_1^2}F_{01}F_{11} + \frac{\sigma^2}{\tau_2^2}F_{02}F_{12} + \frac{\sigma^2}{\tau_1\tau_2}(F_{01}F_{12} + F_{11}F_{02}) = 0,
\end{aligned}$$

$$\begin{aligned}
&\dot{F}_{02} + (1-\gamma)\frac{\alpha}{\tau}\tilde{F}_2 + (1-\gamma)\sigma^2\tilde{F}_2F_{00} + \frac{\alpha}{\tau}F_{02} + \frac{\tilde{F}_5}{\tau_1}F_{01} + \frac{\tilde{F}_6}{\tau_2}F_{02} + \sigma^2F_{00}F_{02} + \frac{\sigma^2}{\tau_1}(F_{00}F_{12} + F_{01}F_{02})\\
&\quad + \frac{\sigma^2}{\tau_2}(F_{00}F_{22} + F_{02}^2) + \frac{\sigma^2}{\tau_1^2}F_{01}F_{12} + \frac{\sigma^2}{\tau_2^2}F_{02}F_{22} + \frac{\sigma^2}{\tau_1\tau_2}(F_{01}F_{22} + F_{02}F_{12}) = 0,
\end{aligned}$$

$$\begin{aligned}
&\dot{F}_{11} - \gamma(1-\gamma)\sigma^2\tilde{F}_1^2 + 2(1-\gamma)\sigma^2\tilde{F}_1F_{01} + \frac{2\tilde{F}_3}{\tau_1}F_{11} + \frac{2\tilde{F}_4}{\tau_2}F_12 + \sigma^2F_{01}^2 + \frac{2\sigma^2}{\tau_1}F_{01}F_{11}\\
&\quad + \frac{2\sigma^2}{\tau_2}F_{01}F_{12} + \frac{\sigma^2}{\tau_1^2}F_{11}^2 + \frac{\sigma^2}{\tau_2^2}F_{12}^2 + \frac{2\sigma^2}{\tau_1\tau_2}F_{11}F_{12} = 0,
\end{aligned}$$

$$\begin{aligned}
&\dot{F}_{12} - \gamma(1-\gamma)\sigma^2\tilde{F}_1\tilde{F}_2 + (1-\gamma)\sigma^2\tilde{F}_2F_{01} + (1-\gamma)\sigma^2\tilde{F}_1F_{02} + \frac{\tilde{F}_5}{\tau_1}F_{11} + \frac{\tilde{F}_3}{\tau_1}F_{12} + \frac{\tilde{F}_6}{\tau_2}F_{12} + \frac{\tilde{F}_4}{\tau_2}F_{22}\\
&\quad + \sigma^2F_{01}F_{02} + \frac{\sigma^2}{\tau_1}(F_{01}F_{12} + F_{02}F_{11}) + \frac{\sigma^2}{\tau_2}(F_{01}F_{22} + F_{02}F_{12}) + \frac{\sigma^2}{\tau_1^2}F_{11}F_{12} + \frac{\sigma^2}{\tau_2^2}F_{12}F_{22}\\
&\quad + \frac{\sigma^2}{\tau_1\tau_2}(F_{11}F_{22} + F_{12}^2) = 0,
\end{aligned}$$

$$\dot{F}_{22} - \gamma(1-\gamma)\sigma^2\tilde{F}_2^2 + 2(1-\gamma)\sigma^2\tilde{F}_2F_{02} + \frac{2\tilde{F}_5}{\tau_1}F_{12} + \frac{2\tilde{F}_6}{\tau_2}F_{22} + \sigma^2F_{02}^2 + \frac{2\sigma^2}{\tau_1}F_{02}F_{12}$$
$$+ \frac{2\sigma^2}{\tau_2}F_{02}F_{22} + \frac{\sigma^2}{\tau_1^2}F_{12}^2 + \frac{\sigma^2}{\tau_2^2}F_{22}^2 + \frac{2\sigma^2}{\tau_1\tau_2}F_{12}F_{22} = 0,$$

$$\dot{F}_0 + (1-\gamma)\frac{\alpha}{\tau}\tilde{F}_7 + \tilde{F}_9F_{00} + \frac{\alpha}{\tau}F_0 + \frac{\tilde{F}_8}{\tau_1}F_{01} + \frac{\tilde{F}_8}{\tau_2}F_{02} + \sigma^2F_{00}F_0 + \frac{\sigma^2}{\tau_1}(F_{01}F_0 + F_{00}F_1)$$
$$+ \frac{\sigma^2}{\tau_2}(F_{02}F_0 + F_{00}F_2) + \frac{\sigma^2}{\tau_1^2}F_{01}F_1 + \frac{\sigma^2}{\tau_2^2}F_{02}F_2 + \frac{\sigma^2}{\tau_1\tau_2}(F_{01}F_2 + F_{02}F_1) = 0,$$

$$\dot{F}_1 + (1-\gamma)\tilde{F}_1\tilde{F}_{10} - \gamma(1-\gamma)\sigma^2\tilde{F}_1\tilde{F}_7 + \tilde{F}_9F_{01} + (1-\gamma)\sigma^2\tilde{F}_1F_0 + \frac{\tilde{F}_8}{\tau_1}F_{11} + \frac{\tilde{F}_8}{\tau_2}F_{12}$$
$$+ \frac{\tilde{F}_3}{\tau_1}F_1 + \frac{\tilde{F}_4}{\tau_2}F_2 + \sigma^2F_{01}F_0 + \frac{\sigma^2}{\tau_1}(F_{01}F_1 + F_{11}F_0) + \frac{\sigma^2}{\tau_2}(F_{01}F_2 + F_{12}F_0) + \frac{\sigma^2}{\tau_1^2}F_{11}F_1$$
$$+ \frac{\sigma^2}{\tau_2^2}F_{11}F_2 + \frac{\sigma^2}{\tau_1\tau_2}(F_{11}F_2 + F_{12}F_1) = 0,$$

$$\dot{F}_2 + (1-\gamma)\tilde{F}_2\tilde{F}_{10} - \gamma(1-\gamma)\sigma^2\tilde{F}_2\tilde{F}_7 + \tilde{F}_9F_{02} + (1-\gamma)\sigma^2\tilde{F}_2F_0 + \frac{\tilde{F}_8}{\tau_1}F_{12} + \frac{\tilde{F}_8}{\tau_2}F_{22}$$
$$+ \frac{\tilde{F}_5}{\tau_1}F_1 + \frac{\tilde{F}_6}{\tau_2}F_2 + \sigma^2F_{02}F_0 + \frac{\sigma^2}{\tau_1}(F_{02}F_1 + F_{12}F_0) + \frac{\sigma^2}{\tau_2}(F_{02}F_2 + F_{22}F_0) + \frac{\sigma^2}{\tau_1^2}F_{12}F_1$$
$$+ \frac{\sigma^2}{\tau_2^2}F_{22}F_2 + \frac{\sigma^2}{\tau_1\tau_2}(F_{12}F_2 + F_{22}F_1) = 0,$$

$$\dot{F}_3 + (1-\gamma)\tilde{F}_{10}\tilde{F}_7 + (1-\gamma)r - \frac{\gamma(1-\gamma)\sigma^2}{2}\tilde{F}_7^2 + \tilde{F}_7F_0 + \frac{\tilde{F}_8}{\tau_1}F_1 + \frac{\tilde{F}_8}{\tau_2}F_2 + \frac{\sigma^2}{\tau}(F_{00} + F_0^2)$$
$$+ \frac{\sigma^2}{\tau_1}(F_{01} + F_0F_1) + \frac{\sigma^2}{\tau_2}(F_{02} + F_0F_2) + \frac{\sigma^2}{2\tau_1^2}(F_{11} + F_1^2) + \frac{\sigma^2}{2\tau_2^2}(F_{22} + F_2^2) + \frac{\sigma^2}{\tau_1\tau_2}(F_{12} + F_1F_2) = 0,$$

with the terminal conditions $F_{ij,T} = F_{i,T} = 0$; additionally, the corresponding coefficients are given by

$$\tilde{F}_{1t} = \frac{\alpha_1}{\gamma\sigma^2} + \frac{A_{11t}}{\tau_1} + \frac{A_{21t}}{\tau_2}, \qquad \tilde{F}_{2t} = \frac{\alpha_2}{\gamma\sigma^2} + \frac{A_{12t}}{\tau_1} + \frac{A_{22t}}{\tau_2},$$
$$\tilde{F}_{3t} = -1 + \alpha_1 + (1-\gamma)\sigma^2\tilde{F}_{1t}, \qquad \tilde{F}_{4t} = \tilde{F}_{3t} + 1,$$
$$\tilde{F}_{5t} = \alpha_2 + (1-\gamma)\sigma^2\tilde{F}_{2t}, \qquad \tilde{F}_{6t} = \tilde{F}_{5t} - 1,$$
$$\tilde{F}_{7t} = \frac{1-\alpha_1-\alpha_2}{\gamma\sigma^2}\mu + \frac{A_{1t}}{\tau_1} + \frac{A_{2t}}{\tau_2}, \qquad \tilde{F}_{8t} = (1-\alpha_1-\alpha_2)\mu + (1-\gamma)\sigma^2\tilde{F}_{7t},$$
$$\tilde{F}_{9t} = -\frac{\alpha}{\tau}s_{t-\tau} + (1-\gamma)\sigma^2\Big(\frac{1-\alpha_1-\alpha_2}{\gamma\sigma^2}\mu + \frac{A_{1t}}{\tau_1} + \frac{A_{2t}}{\tau_2}\Big) + (1-\alpha)\Big(r + \mu - \frac{\sigma^2}{2}\Big),$$
$$\tilde{F}_{10t} = -\frac{\alpha}{\tau}s_{t-\tau} - \alpha r + \frac{\alpha\sigma^2}{2} + (1-\alpha)\mu,$$

where A_{ijt} and A_{it} are deterministic functions given by (II.1) with $N = 2$.

Then, the present value of the CEW is given by

$$\widehat{R} = e^{-rT}\mathrm{CEW}_{\hat{\phi}^N} = e^{-rT}\tilde{f}_0^{\frac{1}{1-\gamma}}$$

$$= \exp\left\{\frac{F_{00,0}s_0^2/2 + \sum_{i=1}^N F_{0i,0}s_t m_{i,0} + \sum_{i=1}^N\sum_{j=1}^N F_{ij,0}m_{i,0}m_{j,0}/2 + F_{0,0}s_0 + \sum_{i=1}^N F_{i,0}m_{i,0} + F_{N+1,0}}{1-\gamma} - rT\right\}.$$

II.3. **Model Calibration for the HA Model** (4.2)**.** In this paper, we focus on the cases of $N = 1, 2$ for the HA model (4.2). Adding more state variables may not be a better fit for the data given the resulting large degrees of freedom, especially when these variables are highly correlated. As shown shortly, the HA model with $N = 1$ seems to be a better fit for the data than the HA model with $N = 2$. (Perhaps paradoxically, the non-Markoivan model (2.1)–(2.2) actually leads to a more tractable characterization of path dependence than a model that uses a large number of Markovian variables. For example, the HA model with $N = 3$ can involve a 15-dimensional system of ODEs as shown in Online Appendix II.1–II.2.)

When $N = 1$, we use one Markovian state variable m_{1t} to approximate momentum. To obtain the time series of m_{1t}, we approximate the integral in (4.2) according to a standard Euler discretization by following Koijen et al. (2009):

$$m_{1t} \approx \sum_{u=1}^{t} e^{-u}R_{t-u+1}, \tag{II.6}$$

where we set $\tau_1 = 1$ and R_t is the excess return of the momentum asset at month t. Then model (4.2) becomes

$$R_{t+1} = (1-\alpha_1)\mu + \alpha_1 m_{1t} + \sigma\epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0,1). \tag{II.7}$$

When $N = 2$, two Markovian state variables m_{1t} and m_{2t} are used to approximate momentum. We set $\tau_1 = 1$ and $\tau_2 = 10$. (We choose significantly different values for τ_1 and τ_2 to make the two independent variables m_{1t} and m_{2t} less correlated in the estimation and hence to better span momentum.) Then model (4.2) becomes

$$R_{t+1} = (1-\alpha_1-\alpha_2)\mu + \alpha_1 m_{1t} + \alpha_2 m_{2t} + \sigma\epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0,1), \tag{II.8}$$

where $m_{1t} \approx \sum_{u=1}^t e^{-u}R_{t-u+1}$ and $m_{2t} \approx \frac{1}{10}\sum_{u=1}^t e^{-u/10}R_{t-u+1}$.

We calibrate models (II.7) and (II.8) based on the size factor using the maximum likelihood method. The estimated parameters are reported in Table II.1. Both coefficients of α_1 and α_2 are positive, and the t-statistics of them are much lower than that for the momentum model (t-statistic $= 2.85$), showing that both m_{1t} and m_{2t} have weaker ability than momentum m_t in predicting future returns. Using the size factor, the Akaike information criteria (AIC) are

-0.8142, -0.8078, and -0.8022 for the momentum model (2.1)–(2.2), the HA model (4.2) with $N = 1$, and the HA model with $N = 2$, respectively. According to the AIC, the momentum model is the best fit for the data, and the HA model with $N = 1$ seems to be a better fit for the data than that with $N = 2$.

Table II.1. Parameter estimations for the HA models

This table reports the estimated parameters of the HA models (II.7) ($N = 1$) and (II.8) ($N = 2$) in annual terms, as well as the related t-statistics (in brackets) based on the factor returns using the maximum likelihood method.

N	α_1	α_2	μ	σ
1	0.17		0.08	0.36
	[1.77]		[4.57]	[36.95]
2	0.03	0.36	0.08	0.36
	[0.24]	[1.95]	[3.29]	[36.80]

Appendix III. Proofs

III.1. **Certainty Equivalent Wealth (CEW).**

III.1.1. *The CEW of the Optimal Momentum Strategy.* The CEW is calculated as $\frac{W_0^{1-\gamma}\text{CEW}^{1-\gamma}}{1-\gamma} = \mathbb{E}_0[\frac{W_T^{1-\gamma}}{1-\gamma}]$. Using (A.3), the CEW of the optimal momentum strategy is given by

$$\text{CEW}_{\phi^*} = \bar{\xi}_0^{\gamma/(1-\gamma)} e^{rT} = e^{rT} \exp\left\{ \frac{\gamma}{1-\gamma}\Big(\frac{A_{1,0}}{2}(\ln S_0)^2 + A_{2,0}\ln S_0 + A_{3,0}\Big)\right\}, \tag{III.1}$$

where the last equality follows from (A.6) and (A.9), and $A_{1,0}$, $A_{2,0}$ and $A_{3,0}$ are given by (A.13) in Appendix A.2. In this paper, we study the present value of CEW_{ϕ^*}, which is defined as $R^* = e^{-rT}\text{CEW}_{\phi^*}$.

III.1.2. *The CEW of the Mean-Variance Strategy.* The wealth process of an investor who uses the mean-variance strategy follows $dW_t/W_t = (r + \theta_t^2/\gamma)dt + \theta_t/\gamma dB_t$. The corresponding value function is given by

$$\begin{aligned} J_t^M &= \mathbb{E}_t\Big[\frac{W_T^{1-\gamma}}{1-\gamma}\Big] = e^{(1-\gamma)r(T-t)}\frac{W_t^{1-\gamma}}{1-\gamma}\mathbb{E}_t\Big[\exp\Big\{\int_t^T \frac{(2\gamma-1)(1-\gamma)}{2\gamma^2}\theta_u^2 du + \frac{1-\gamma}{\gamma}\theta_u dB_u\Big\}\Big] \\ &= e^{(1-\gamma)r(T-t)}\frac{W_t^{1-\gamma}}{1-\gamma}\mathbb{E}_t^*\Big[\exp\Big\{\frac{1-\gamma}{2\gamma\sigma^2}\int_t^T \Big[(1-\alpha)\mu - \alpha\Big(r - \frac{\sigma^2}{2}\Big) + \frac{\alpha}{\tau}(s_u - s_{u-\tau})\Big]^2 du\Big\}\Big], \end{aligned} \tag{III.2}$$

where the last equality follows the change of measure in Appendix A.2. The Feynman-Kac formula implies that

$$\begin{aligned}&\frac{\partial J^M}{\partial t}+\Big[\Big(r-\frac{\sigma^2}{2}\Big)\Big(1-\frac{\alpha}{\gamma}\Big)+(1-\alpha)\frac{\mu}{\gamma}+\frac{\alpha}{\gamma\tau}(s-s_{t-\tau})\Big]\frac{\partial J^M}{\partial s}+\frac{\sigma^2}{2}\frac{\partial^2 J^M}{\partial s^2}\\&+\frac{1-\gamma}{2\gamma\sigma^2}\Big[(1-\alpha)\mu-\alpha\Big(r-\frac{\sigma^2}{2}\Big)+\frac{\alpha}{\tau}(s-s_{t-\tau})\Big]^2 J^M=0.\end{aligned}\tag{III.3}$$

The solution to (III.3) is given by $J^M(s,t)=\exp\{A^M_{1,t}s_t^2/2+A^M_{2,t}s_t+A^M_{3,t}\}$, where $A^M_{1,t}$, $A^M_{2,t}$ and $A^M_{3,t}$ are governed by the ODEs

$$\begin{aligned}\dot{A}^M_{1,t}&=-\sigma^2(A^M_{1,t})^2-\frac{2\alpha}{\gamma\tau}A^M_{1,t}-\frac{1-\gamma}{\gamma}\frac{\alpha^2}{\tau^2\sigma^2},\\ \dot{A}^M_{2,t}&=-\Big(\sigma^2A^M_{1,t}+\frac{\alpha}{\gamma\tau}\Big)A^M_{2,t}-\Big[(1-\alpha)\frac{\mu}{\gamma}+\Big(r-\frac{\sigma^2}{2}\Big)\Big(1-\frac{\alpha}{\gamma}\Big)-\frac{\alpha}{\gamma\tau}s_{t-\tau}\Big]A^M_{1,t}\\&\quad-\frac{1-\gamma}{\gamma}\frac{\alpha}{\sigma^2\tau}\Big[(1-\alpha)\mu+\alpha\Big(\frac{\sigma^2}{2}-r\Big)-\frac{\alpha}{\tau}s_{t-\tau}\Big],\\ \dot{A}^M_{3,t}&=-\frac{\sigma^2}{2}(A^M_{2,t})^2-\frac{\sigma^2}{2}A^M_{1,t}-\Big[(1-\alpha)\frac{\mu}{\gamma}+\Big(r-\frac{\sigma^2}{2}\Big)\Big(1-\frac{\alpha}{\gamma}\Big)-\frac{\alpha}{\gamma\tau}s_{t-\tau}\Big]A^M_{2,t}\\&\quad-\frac{1-\gamma}{2\gamma\sigma^2}\Big[(1-\alpha)\mu+\alpha\Big(\frac{\sigma^2}{2}-r\Big)-\frac{\alpha}{\tau}s_{t-\tau}\Big]^2,\end{aligned}$$

which has the terminal conditions $A^M_{1,T}=A^M_{2,T}=A^M_{3,T}=0$. Thus, the present value of the CEW is given by

$$R^M=(J^M_0)^{1/(1-\gamma)}=\exp\Big\{\Big(\frac{A^M_{1,0}s_0^2}{2}+A^M_{2,0}s_0+A^M_{3,0}\Big)/(1-\gamma)\Big\}.$$

III.2. **Proof of Proposition 6.1.** The market is complete, and the Cox-Huang approach applies. The market prices of risk, namely, $\boldsymbol{\theta}_t$, are given by (6.4), and the state price density is given by

$$\pi_t=\exp\Big\{-\int_0^t r\,du-\frac{1}{2}\int_0^t\boldsymbol{\theta}_u'\boldsymbol{\theta}_u du-\int_0^t\boldsymbol{\theta}_u' d\mathbf{B}_u\Big\},\tag{III.4}$$

where

$$d\mathbf{B}_t=\begin{pmatrix}d\mathbf{B}^F_t\\ d\mathbf{B}^M_t\end{pmatrix},\qquad d\mathbf{B}^F_t=(dB^F_{1,t},dB^F_{2,t},\cdots,dB^F_{N,t})'.\tag{III.5}$$

Define

$$\xi_t=\exp\Big\{-\frac{1}{2}\int_0^t\boldsymbol{\theta}_u'\boldsymbol{\theta}_u du-\int_0^t\boldsymbol{\theta}_u' d\mathbf{B}_u\Big\}.\tag{III.6}$$

We rewrite the momentum as

$$\mathbf{m}_t=\frac{1}{\tau}(\mathbf{s}_t-\mathbf{s}_{t-\tau})-r+\frac{1}{2}(\boldsymbol{\Sigma}^2\mathbf{1}+\mathbf{v}),\tag{III.7}$$

where $\mathbf{s}_t = (\ln S_{1,t}, \ln S_{2,t}, \cdots, \ln S_{N,t})'$ is a vector of log prices, $\mathbf{1}$ is an $N \times 1$ vector of 1, and $\mathbf{v}$ is an $N \times 1$ vector whose ith element is given by the iith element of $\boldsymbol{\beta}\boldsymbol{\beta}'$. Then, the stocks' log price follows

$$d\mathbf{s}_t = \Big[a + \frac{b}{\tau}(\mathbf{s}_t - \mathbf{s}_{t-\tau}) + \boldsymbol{\beta}\boldsymbol{\alpha} + (1-b)r - \frac{1-b}{2}(\boldsymbol{\Sigma}^2\mathbf{1} + \mathbf{v})\Big]dt + \boldsymbol{\beta}d\mathbf{B}_t^F + \boldsymbol{\Sigma}d\mathbf{B}_t^M. \quad \text{(III.8)}$$

We define a new measure:

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = \exp\Big\{ - \int_t^T \frac{\gamma-1}{\gamma}\boldsymbol{\theta}_u' d\mathbf{B}_u - \int_t^T \frac{(\gamma-1)^2}{2\gamma^2}\boldsymbol{\theta}_u'\boldsymbol{\theta}_u du\Big\}. \quad \text{(III.9)}$$

Under this measure,

$$\begin{aligned} d\mathbf{s}_t =& \Big[a + \frac{b}{\tau}(\mathbf{s}_t - \mathbf{s}_{t-\tau}) + \boldsymbol{\beta}\boldsymbol{\alpha} + (1-b)r - \frac{1-b}{2}(\boldsymbol{\Sigma}^2\mathbf{1} + \mathbf{v}) - \frac{\gamma-1}{\gamma}\begin{pmatrix} \boldsymbol{\beta} & \boldsymbol{\Sigma} \end{pmatrix}\boldsymbol{\theta}\Big]dt \\ &+ \boldsymbol{\beta}d\mathbf{B}_t^{F*} + \boldsymbol{\Sigma}d\mathbf{B}_t^{M*}. \end{aligned} \quad \text{(III.10)}$$

where

$$\begin{aligned} \boldsymbol{\theta} =& \boldsymbol{\Omega}^{-1}\boldsymbol{\mu} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\beta}' \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{\alpha} - r \\ a + \frac{b}{\tau}(\mathbf{s}_t - \mathbf{s}_{t-\tau}) - br + \frac{b}{2}(\boldsymbol{\Sigma}^2\mathbf{1} + \mathbf{v}) \end{pmatrix} \\ =& \begin{pmatrix} \boldsymbol{\alpha} - r - \boldsymbol{\beta}'\Big[a + \frac{b}{\tau}(\mathbf{s}_t - \mathbf{s}_{t-\tau}) - br + \frac{b}{2}(\boldsymbol{\Sigma}^2\mathbf{1} + \mathbf{v})\Big] \\ \boldsymbol{\Sigma}^{-1}\Big[a + \frac{b}{\tau}(\mathbf{s}_t - \mathbf{s}_{t-\tau}) - br + \frac{b}{2}(\boldsymbol{\Sigma}^2\mathbf{1} + \mathbf{v})\Big] \end{pmatrix}. \end{aligned} \quad \text{(III.11)}$$

Denote

$$f(\mathbf{s}, t) = \mathbb{E}_t[\xi_T^{\frac{\gamma-1}{\gamma}}] = \mathbb{E}_t^*\Big[\exp\Big\{\frac{1-\gamma}{2\gamma^2}\int_t^T \boldsymbol{\theta}_u'\boldsymbol{\theta}_u du\Big\}\Big]. \quad \text{(III.12)}$$

When $t \leq T$, the Feynman-Kac formula leads to

$$\begin{aligned} &\frac{\partial f}{\partial t} + \Big[a + \frac{b}{\tau}(\mathbf{s}_t - \mathbf{s}_{t-\tau}) + \boldsymbol{\beta}\boldsymbol{\alpha} + (1-b)r - \frac{1-b}{2}(\boldsymbol{\Sigma}^2\mathbf{1} + \mathbf{v}) - \frac{\gamma-1}{\gamma}\begin{pmatrix} \boldsymbol{\beta} & \boldsymbol{\Sigma} \end{pmatrix}\boldsymbol{\theta}\Big]'\frac{\partial f}{\partial \mathbf{s}} \\ &+ \frac{1}{2}tr\Big(\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\frac{\partial^2 f}{\partial \mathbf{s}^2}\Big) + \frac{1-\gamma}{2\gamma^2}\boldsymbol{\theta}'\boldsymbol{\theta}f = 0. \end{aligned} \quad \text{(III.13)}$$

Following the proof in Appendix A.2, the optimal portfolio weights are given by

$$\boldsymbol{\phi}^* = \frac{1}{\gamma}(\boldsymbol{\Omega}\boldsymbol{\Omega}')^{-1}\begin{pmatrix} \boldsymbol{\alpha} - r \\ a + b\mathbf{m} + \boldsymbol{\beta}\boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\phi}^H \end{pmatrix}, \quad \text{(III.14)}$$

where

$$(\boldsymbol{\Omega}\boldsymbol{\Omega}')^{-1} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\beta}' \\ -\boldsymbol{\beta} & \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}^{-2} \end{pmatrix}. \quad \text{(III.15)}$$

The hedging demands for the factor assets are zero because their returns are IID. The hedging demands for the momentum assets are given by

$$\boldsymbol{\phi}^H = (1-\gamma)(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}^{-2}) \int_{-\tau}^{0} \omega_v (d\mathbf{R}_v - \mathbf{1}r dv) + \mathbf{C}, \tag{III.16}$$

and

$$\begin{aligned}
\omega_v &= \begin{cases} \int_{-\tau}^{v} \hat{\omega}_u du, & v \in [-\tau, -\tau+T], \\ \int_{-\tau}^{-\tau+T} \hat{\omega}_u du, & v \in [-\tau+T, 0], \end{cases} \\
\mathbf{C} &= \int_0^T \exp\left\{ \int_0^u \Big[\frac{\gamma(1-\gamma)\tau}{b}\hat{C}_{\hat{u}} + \frac{b}{\tau}\Big] d\hat{u} \right\} \left\{ \Big[\frac{a + \boldsymbol{\beta}\boldsymbol{\alpha} - br}{\gamma} + r\Big] \frac{\gamma(1-\gamma)\tau}{b} (\boldsymbol{\Sigma}^{-2}\mathbf{1} + \mathbf{v})\hat{C}_u \right. \\
&\quad \left. - \frac{\gamma(1-\gamma)\tau}{2b}\hat{C}_u + \frac{(\gamma-1)br}{\gamma\tau}(\boldsymbol{\Sigma}^{-2}\mathbf{1} + \mathbf{v}) - \frac{(\gamma-1)b}{2\gamma\tau} \right\} du + \frac{(1-\gamma)br}{\gamma}(\boldsymbol{\Sigma}^{-2}\mathbf{1} + \mathbf{v}) \int_{-\tau}^{0} \omega_v dv,
\end{aligned} \tag{III.17}$$

where

$$\begin{aligned}
\hat{\omega}_u &= \hat{C}_{u+\tau} \exp\left\{ \int_0^{u+\tau} \Big[\frac{\gamma(1-\gamma)\tau}{b}\hat{C}_{\hat{u}} + \frac{b}{\tau}\Big] d\hat{u} \right\} > 0, \\
\hat{C}_u &= \frac{b^2\big(e^{\frac{2b(T-u)}{\sqrt{\gamma}\tau}} + 1\big)}{\gamma^{3/2}\tau^2\Big[(\sqrt{\gamma}-1)e^{\frac{2b(T-u)}{\sqrt{\gamma}\tau}} + (\sqrt{\gamma}+1)\Big]}.
\end{aligned} \tag{III.18}$$

III.3. **Proof of Proposition 6.2.** The wealth process follows

$$dW_t = W_t\Big\{ r + \phi_t^R \sigma_t \Big[\frac{\alpha}{\tau}(R_t - R_{t-\tau}) + (1-\alpha)\mu^R\Big]\Big\} dt + W_t \phi_t^R \sigma_t dB_t, \tag{III.19}$$

where ϕ_t^R is the portfolio weight of the examined risky asset. When $T \leq \tau$, the investor's value function, $J(t, W, R)$, is governed by (Li and Liu, 2018):

$$\begin{aligned}
\max_{\phi^R} \left\{ \frac{\partial J}{\partial t} + W\Big\{ r + \phi^R \sigma \Big[\frac{\alpha}{\tau}(R - R_{t-\tau}) + (1-\alpha)\mu^R\Big]\Big\} \frac{\partial J}{\partial W} + \frac{W^2 \phi^{R2}\sigma^2}{2}\frac{\partial^2 J}{\partial W^2} \right. \\
\left. + \Big[\frac{\alpha}{\tau}(R - R_{t-\tau}) + (1-\alpha)\mu^R\Big]\frac{\partial J}{\partial R} + W\phi^R\sigma\frac{\partial^2 J}{\partial W \partial R} + \frac{1}{2}\frac{\partial^2 J}{\partial R^2} \right\} = 0,
\end{aligned} \tag{III.20}$$

with the boundary condition $J(T, W, R) = \frac{W^{1-\gamma}}{1-\gamma}$. By conjecturing that $J = \frac{W^{1-\gamma}}{1-\gamma}[f(R,t)]^\gamma$ and using FOC, we obtain

$$\frac{\partial f}{\partial t} + \frac{1-\gamma}{\gamma} r f + \frac{1-\gamma}{2\gamma^2}\Big[\frac{\alpha}{\tau}(R - R_{t-\tau}) + (1-\alpha)\mu^R\Big]^2 f + \frac{1}{\gamma}\Big[\frac{\alpha}{\tau}(R - R_{t-\tau}) + (1-\alpha)\mu^R\Big]\frac{\partial f}{\partial R} + \frac{1}{2}\frac{\partial^2 f}{\partial R^2} = 0,$$

where $f(R,T)=1$. The solution is given by $f(R,t)=\exp\{A_1(t)R^2/2+A_2(t)R+A_3(t)\}$, where

$$A_1(t)=\frac{\frac{\gamma-1}{\gamma}\frac{\alpha}{\tau}\tanh[\frac{-\alpha}{\sqrt{\gamma\tau}}(T-t)]}{\sqrt{\gamma}+\tanh[\frac{-\alpha}{\sqrt{\gamma\tau}}(T-t)]},$$

$$A_2(t)=\int_t^T\Big[(1-\alpha)\mu^R-\frac{\alpha}{\tau}R_{s-\tau}\Big]\Big(\frac{1-\gamma}{\gamma^2}\frac{\alpha}{\tau}+\frac{A_1(s)}{\gamma}\Big)e^{\int_t^s(\frac{1}{\gamma}\frac{\alpha}{\tau}+A_1(v))dv}ds,$$

$$A_3(t)=\int_t^T\frac{1-\gamma}{\gamma}r+\frac{1-\gamma}{2\gamma^2}\Big[(1-\alpha)\mu^R-\frac{\alpha}{\tau}R_{s-\tau}\Big]^2+\frac{A_2(s)}{\gamma}\Big[(1-\alpha)\mu^R-\frac{\alpha}{\tau}R_{s-\tau}\Big]+\frac{A_1(s)+A_2^2(s)}{2}ds.$$

The optimal portfolio weight of the examined risky asset is given by

$$\begin{aligned}\phi^{R*}&=\frac{1}{\gamma\sigma_t}\Big\{\Big[\alpha m_t^R+(1-\alpha)\mu^R\Big]+\gamma[A_1(t)R_t+A_2(t)]\Big\}\\&=\frac{\alpha m_t^R+(1-\alpha)\mu^R}{\gamma\sigma_t}+\frac{1-\gamma}{\sigma_t}\Big[\int_{t-\tau}^t\omega^R(v,t)dR_v+C^R(t)\Big],\end{aligned}$$

where

$$\omega^R(v,t)=\begin{cases}\int_t^{v+\tau}\hat{\omega}^R(u,t)du, & v\in[t-\tau,T-\tau],\\ \int_t^T\hat{\omega}^R(u,t)du, & v\in[T-\tau,t],\end{cases}\tag{III.21}$$

$$C^R(t)=(1-\alpha)\mu^R\int_t^T\Big[\frac{1-\gamma}{\gamma^2}\frac{\alpha}{\tau}+\frac{A_1(u)}{\gamma}\Big]e^{\int_t^u[\frac{1}{\gamma}\frac{\alpha}{\tau}+A_1(v)]dv}du,$$

where $\hat{\omega}^R(u,t)=\frac{\alpha}{\gamma\tau}\big[\frac{\alpha}{\gamma\tau}+\frac{A_1(u)}{1-\gamma}\big]e^{\int_t^u[\frac{1}{\gamma}\frac{\alpha}{\tau}+A_1(v)]dv}$.

III.4. **Proof of Proposition 6.3.** When $0\le T-t\le\tau$, $s_{u-\tau}$ for all $u\le T$ are realized (log) prices and are known at time t ($\le T$). In this case, the standard dynamic programming approach applies. Let $J(t,W,s)$ denote the indirect utility function. Bellman's principle of optimality leads to the following Hamilton-Jacobi-Bellman (HJB) equation (Merton, 1971) for J:

$$\begin{aligned}\max_\phi\Big\{&\frac{\partial J}{\partial t}+\frac{\partial J}{\partial W}W\Big[r+\Big[\frac{\alpha(s-s_{t-\tau})}{\tau}-\alpha r+\frac{\alpha\sigma^2}{2}+(1-\alpha)\mu\Big]\phi\Big]+\frac{\partial^2 J}{\partial W^2}\frac{\sigma^2W^2\phi^2}{2}\\&+\frac{\partial J}{\partial s}\Big[\frac{\alpha(s-s_{t-\tau})}{\tau}+(1-\alpha)\Big(r+\mu-\frac{\sigma^2}{2}\Big)\Big]+\frac{\partial^2 J}{\partial W\partial s}\sigma^2W\phi+\frac{\partial^2 J}{\partial s^2}\frac{\sigma^2}{2}\Big\}=0,\end{aligned}\tag{III.22}$$

with the boundary condition $J(T,W,s)=\frac{W^{1-\gamma}}{1-\gamma}$.

The solution for (III.22) is given by

$$J(t,W,s)=\frac{W^{1-\gamma}}{1-\gamma}\exp\Big\{\gamma\Big(\frac{A_{11t}^{(1)}}{2}s^2+A_{1t}^{(1)}s+A_{2t}^{(1)}\Big)\Big\},\tag{III.23}$$

where

$$
\begin{aligned}
A_{11t}^{(1)} &= \frac{-\gamma g_1 \tanh[\frac{-\alpha}{\sqrt{\gamma\tau}}(T-t)]}{\sqrt{\gamma}+\tanh[\frac{-\alpha}{\sqrt{\gamma\tau}}(T-t)]},\\
A_{1t}^{(1)} &= \int_t^T \Big[g_1\Big(g_3-\frac{\alpha}{\tau}s_{u-\tau}\Big)+\Big(g_2-\frac{\alpha}{\gamma\tau}s_{u-\tau}\Big)A_{11u}^{(1)}\Big]e^{\int_t^u \frac{\alpha}{\gamma\tau}+\sigma^2 A_{11s}^{(1)}ds}du, \qquad \text{(III.24)}\\
A_{2t}^{(1)} &= \int_t^T g_4+\frac{\tau g_1}{2\alpha}\Big(g_3-\frac{\alpha}{\tau}s_{u-\tau}\Big)^2+\Big(g_2-\frac{\alpha}{\gamma\tau}s_{u-\tau}\Big)A_{1u}^{(1)}+\frac{\sigma^2}{2}(A_{1u}^{(1)})^2+\frac{\sigma^2}{2}A_{11u}^{(1)}du,
\end{aligned}
$$

for $0 \le t \le T$, $g_1 = \frac{1-\gamma}{\gamma^2}\frac{\alpha}{\sigma^2\tau}$, $g_2 = \frac{1-\alpha}{\gamma}\mu - \frac{\gamma-\alpha}{\gamma}\frac{\sigma^2}{2} + \frac{\gamma-\alpha}{\gamma}r$, $g_3 = (1-\alpha)\mu + \frac{\alpha\sigma^2}{2} - \alpha r$, and $g_4 = \frac{1-\gamma}{\gamma}r$ are constants. Note that $A_{1t}^{(1)}$ and $A_{2t}^{(1)}$ depend on the historical price path s_u for $u \in [t-\tau, T-\tau]$. The optimal portfolio weight is given by $\phi_t^* = \phi_t^M + \phi_t^{sH} + \phi_t^{pH}$, where the myopic demand is given by

$$
\phi_t^M = \frac{\alpha m_t + (1-\alpha)\mu}{\gamma\sigma^2}; \qquad \text{(III.25)}
$$

additionally, the "momentum hedging demand" and "path hedging demand" are given, respectively, by

$$
\phi_t^{sH} = A_{11t}^{(1)} s_t \quad \text{and} \quad \phi_t^{pH} = A_{1t}^{(1)},
$$

which are caused by time-varying momentum and price paths, respectively.

When $T-t \ge \tau$, the delayed variable $s_{u-\tau}$ becomes unknown at time t for $u > t+\tau$, and it cannot be characterized by s_u. As a result, the standard dynamic programming approach cannot be applied. In this case, more state variables are needed to constitute a sufficient statistic of the indirect utility function. We show that these state variables are different for different investment horizons. In fact, as the horizon lengthens, the number of state variables required increases without bound. Therefore, we can only solve this problem piecewise. For a given investment horizon, we will first introduce some path-induced state variables that together with the original state variables, constitute a sufficient statistic. Then, we can write down the HJB equation with respect to all the state variables.

Now, we study the case $\tau < T-t \le 2\tau$, which we use to highlight the construction process of the new state variables. When $0 \le t \le T-\tau$,

$$
\begin{aligned}
J_t &= \max_\phi \big\{\mathbb{E}_t[J_T]\big\} = \max_\phi \big\{\mathbb{E}_t[J_{T-\tau}]\big\}\\
&= \max_\phi \mathbb{E}_t\Big\{\frac{W_{T-\tau}^{1-\gamma}}{1-\gamma}\exp\Big\{\gamma\Big(\frac{A_{11T-\tau}^{(1)}}{2}s_{T-\tau}^2 + A_{1T-\tau}^{(1)}s_{T-\tau} + A_{2T-\tau}^{(1)}\Big)\Big\}\Big\}.
\end{aligned} \qquad \text{(III.26)}
$$

The first two equalities of (III.26) follow Bellman's principle of optimality, and the last equality follows (III.23). Note that $A_{11T-\tau}^{(1)}$ is a (deterministic) constant, but $A_{1T-\tau}^{(1)}$ and $A_{2T-\tau}^{(1)}$ are stochastic and are not $\mathcal{F}_t$-measurable. In fact, $A_{1T-\tau}^{(1)}$ and $A_{2T-\tau}^{(1)}$ are governed by backward random ODEs, so they depend on all the s_u of $u \in [T-2\tau, T-\tau]$, which are not completely known at time t. Thus, they lead to new state variables. In the following analysis, we first rewrite the last equation of (III.26) by introducing new adapted state variables that are forward, then we write the HJB equation based on these new state variables, along with the original state variables s and W. Finally, we derive the solutions of the HJB equation.

$A_1^{(1)}$ in (III.24) can be rewritten as

$$A_{1t}^{(1)} = \int_{t-\tau}^{T-\tau} \Big[g_1\Big(g_3 - \frac{\alpha}{\tau}s_u\Big) + \Big(g_2 - \frac{\alpha}{\gamma\tau}s_u\Big)A_{11u+\tau}^{(1)}\Big] e^{\int_t^{u+\tau} \frac{\alpha}{\gamma\tau} + \sigma^2 A_{11s}^{(1)} ds} du. \tag{III.27}$$

We define a new variable, namely,

$$\tilde{B}_{2t}^{(2)} = \int_{T-2\tau}^{t} \Big[g_1\Big(g_3 - \frac{\alpha}{\tau}s_u\Big) + \Big(g_2 - \frac{\alpha}{\gamma\tau}s_u\Big)A_{11u+\tau}^{(1)}\Big] f_u^{(1)} du, \tag{III.28}$$

for $T-2\tau \le t \le T-\tau$, where $f_u^{(1)} = e^{\int_{T-\tau}^{u+\tau} \frac{\alpha}{\gamma\tau} + \sigma^2 A_{11s}^{(1)} ds}$ is a deterministic function. Note that $\tilde{B}_{2t}^{(2)}$ is forward, adapted to $\mathcal{F}_t$, and governed by a random ODE:

$$\dot{\tilde{B}}_{2t}^{(2)} = \Big[g_1\Big(g_3 - \frac{\alpha}{\tau}s_t\Big) + \Big(g_2 - \frac{\alpha}{\gamma\tau}s_t\Big)A_{11t+\tau}^{(1)}\Big] f_t^{(1)}, \quad \text{with} \quad \tilde{B}_{T-2\tau}^{(2)} = 0. \tag{III.29}$$

Then, we can express $A_1^{(1)}$ as $\tilde{B}_2^{(2)}$:

$$A_{1t}^{(1)} = (f_{t-\tau}^{(1)})^{-1}(\tilde{B}_{2T-\tau}^{(2)} - \tilde{B}_{2t-\tau}^{(2)}). \tag{III.30}$$

Similarly, we can transform $A_{2T-\tau}^{(1)}$ in (III.26) into new state variables. By substituting (III.30) into the last equation of (III.24), we obtain

$$A_{2T-\tau}^{(1)} = g_0(\tilde{B}_{2T-\tau}^{(2)})^2 + \tilde{B}_{2T-\tau}^{(2)}\tilde{B}_{3T-\tau}^{(2)} + \tilde{B}_{4T-\tau}^{(2)}, \tag{III.31}$$

where $g_0 = \int_{T-2\tau}^{T-\tau} \frac{\sigma^2}{2}(f_u^{(1)})^2 du$ is a constant, and $\tilde{B}_3^{(2)}$ and $\tilde{B}_3^{(2)}$ are governed by

$$\begin{aligned} \dot{\tilde{B}}_{3t}^{(2)} &= \Big(g_2 - \frac{\alpha}{\gamma\tau}s_t\Big)f_t^{(1)} - \sigma^2(f_t^{(1)})^2\tilde{B}_{2t}^{(2)}, \\ \dot{\tilde{B}}_{4t}^{(2)} &= g_4 + \frac{\tau g_1}{2\alpha}\Big(g_3 - \frac{\alpha}{\tau}s_t\Big)^2 - \Big(g_2 - \frac{\alpha}{\gamma\tau}s_t\Big)f_t^{(1)}\tilde{B}_{2t}^{(2)} + \frac{\sigma^2}{2}A_{11t+\tau}^{(1)} + \frac{\sigma^2}{2}(f_t^{(1)})^2(\tilde{B}_{2t}^{(2)})^2, \end{aligned} \tag{III.32}$$

for $T-2\tau \le t \le T-\tau$, with $\tilde{B}_{3T-2\tau}^{(2)} = \tilde{B}_{4T-2\tau}^{(2)} = 0$. Note that $A_{2T-\tau}^{(1)}$ leads to the introduction of two more state variables due to the coefficient of $\tilde{B}_{2T-\tau}^{(2)}$ and the term that is independent of $\tilde{B}_{2T-\tau}^{(2)}$ in (III.31).

The new state variables $\tilde{B}_i^{(2)}$ for $i = 2, 3, 4$, which are induced by price paths, together with the original state variables s and W, constitute a sufficient statistic of the indirect utility function. The last equation of (III.26) can be rewritten in terms of all the state variables as

$$
\begin{aligned}
&J_t = J(t, W, s, \tilde{B}_2^{(2)}, \tilde{B}_3^{(2)}, \tilde{B}_4^{(2)}) \\
&= \max_{\phi} \mathbb{E}_t \left\{ \frac{W_{T-\tau}^{1-\gamma}}{1-\gamma} \exp\left\{ \gamma \Big(\frac{A_{11T-\tau}^{(1)}}{2} s_{T-\tau}^2 + \tilde{B}_{2T-\tau}^{(2)} s_{T-\tau} + g_0 (\tilde{B}_{2T-\tau}^{(2)})^2 + \tilde{B}_{2T-\tau}^{(2)} \tilde{B}_{3T-\tau}^{(2)} + \tilde{B}_{4T-\tau}^{(2)} \Big) \right\} \right\}.
\end{aligned}
$$

Then, the HJB equation is given by

$$
\begin{aligned}
\max_{\phi} \Bigg\{ & \frac{\partial J}{\partial t} + \frac{\partial J}{\partial W} W \Big[r + \Big[\frac{\alpha(s - s_{t-\tau})}{\tau} - \alpha r + \frac{\alpha \sigma^2}{2} + (1-\alpha)\mu \Big] \phi \Big] + \frac{\partial^2 J}{\partial W^2} \frac{\sigma^2 W^2 \phi^2}{2} \\
& + \frac{\partial J}{\partial s} \Big[\frac{\alpha(s - s_{t-\tau})}{\tau} + (1-\alpha) \Big(r + \mu - \frac{\sigma^2}{2} \Big) \Big] + \frac{\partial^2 J}{\partial W \partial s} \sigma^2 W \phi + \frac{\partial^2 J}{\partial s^2} \frac{\sigma^2}{2} \\
& + \frac{\partial J}{\partial \tilde{B}_2^{(2)}} \Big[g_1 \Big(g_3 - \frac{\alpha}{\tau} s \Big) + \Big(g_2 - \frac{\alpha}{\gamma\tau} s \Big) A_{11t+\tau}^{(1)} \Big] f^{(1)} + \frac{\partial J}{\partial \tilde{B}_3^{(2)}} \Big[\Big(g_2 - \frac{\alpha}{\gamma\tau} s \Big) f^{(1)} - \sigma^2 (f^{(1)})^2 \tilde{B}_2^{(2)} \Big] \\
& + \frac{\partial J}{\partial \tilde{B}_4^{(2)}} \Big[g_4 + \frac{\tau g_1}{2\alpha} \Big(g_3 - \frac{\alpha}{\tau} s \Big)^2 - \Big(g_2 - \frac{\alpha}{\gamma\tau} s_t \Big) f^{(1)} \tilde{B}_2^{(2)} + \frac{\sigma^2}{2} A_{11t+\tau}^{(1)} + \frac{\sigma^2}{2} (f^{(1)})^2 (\tilde{B}_2^{(2)})^2 \Big] \Bigg\} = 0,
\end{aligned}
$$

with the boundary condition $J_{T-\tau} = \frac{W_{T-\tau}^{1-\gamma}}{1-\gamma} \exp\{\gamma(\frac{A_{11T-\tau}^{(1)}}{2} s_{T-\tau}^2 + \tilde{B}_{2T-\tau}^{(2)} s_{T-\tau} + g_0(\tilde{B}_{2T-\tau}^{(2)})^2 + \tilde{B}_{2T-\tau}^{(2)} \tilde{B}_{3T-\tau}^{(2)} + \tilde{B}_{4T-\tau}^{(2)})\}$. Its solution is given by

$$
J(t, W, s, \tilde{B}_2^{(2)}, \tilde{B}_3^{(2)}, \tilde{B}_4^{(2)}) = \frac{W_t^{1-\gamma}}{1-\gamma} \exp\left\{ \gamma \Big(\frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} A_{ijt}^{(2)} \tilde{B}_{it}^{(2)} \tilde{B}_{jt}^{(2)} + \sum_{i=1}^{3} A_{it}^{(2)} \tilde{B}_{it}^{(2)} + \tilde{B}_{4t}^{(2)} + A_{4t}^{(2)} \Big) \right\},
$$

where $\tilde{B}_{1t}^{(2)} \equiv s_t$; the coefficient $A^{(2)}$ is given by

$$\dot{A}_{11t}^{(2)} + \frac{2\alpha g_1}{\tau} + \frac{2\alpha}{\gamma\tau} A_{11t}^{(2)} + \sigma^2 (A_{11t}^{(2)})^2 - \frac{2\alpha}{\tau} f_t^{(1)} \Big(g_1 + \frac{A_{11t+\tau}^{(1)}}{\gamma} \Big) A_{12t}^{(2)} - \frac{2\alpha}{\gamma\tau} f_t^{(1)} A_{13t}^{(2)} = 0,$$

$$\dot{A}_{22t}^{(2)} + \sigma^2 (A_{12t}^{(2)})^2 - 2\sigma^2 (f_t^{(1)})^2 A_{23t}^{(2)} + \sigma^2 (f_t^{(1)})^2 = 0, \qquad \dot{A}_{33t}^{(2)} + \sigma^2 (A_{13t}^{(2)})^2 = 0,$$

$$\dot{A}_{12t}^{(2)} + \frac{\alpha}{\gamma\tau} A_{12t}^{(2)} + \sigma^2 A_{11t}^{(2)} A_{12t}^{(2)} - f_t^{(1)} \frac{\alpha}{\tau} \Big(g_1 + \frac{A_{11t+\tau}^{(1)}}{\gamma} \Big) A_{22t}^{(2)} - \frac{\alpha}{\gamma\tau} f_t^{(1)} A_{23t}^{(2)} - \sigma^2 (f_t^{(1)})^2 A_{13t}^{(2)} + \frac{\alpha}{\gamma\tau} f_t^{(1)} = 0,$$

$$\dot{A}_{13t}^{(2)} + \frac{\alpha}{\gamma\tau} A_{13t}^{(2)} + \sigma^2 A_{11t}^{(2)} A_{13t}^{(2)} - f_t^{(1)} \frac{\alpha}{\tau} \Big(g_1 + \frac{A_{11t+\tau}^{(1)}}{\gamma} \Big) A_{23t}^{(2)} - \frac{\alpha}{\gamma\tau} f_t^{(1)} A_{33t}^{(2)} = 0,$$

$$\dot{A}_{23t}^{(2)} + \sigma^2 A_{12t}^{(2)} A_{13t}^{(2)} - \sigma^2 (f_t^{(1)})^2 A_{33t}^{(2)} = 0,$$

$$\dot{A}_{1t}^{(2)} - \frac{\alpha g_1}{\tau} s_{t-\tau} + \Big(g_2 - \frac{\alpha}{\gamma\tau} s_{t-\tau} \Big) A_{11t}^{(2)} + (g_1 g_3 + g_2 A_{11t+\tau}^{(1)}) f_t^{(1)} A_{12t}^{(2)} + g_2 f_t^{(1)} A_{13t}^{(2)} + \sigma^2 A_{11t}^{(2)} A_{1t}^{(2)}$$
$$+ \frac{\alpha}{\gamma\tau} A_{1t}^{(2)} - \frac{\alpha}{\tau} \Big(g_1 + \frac{A_{11t+\tau}^{(1)}}{\gamma} \Big) f_t^{(1)} A_{2t}^{(2)} - \frac{\alpha}{\gamma\tau} f_t^{(1)} A_{3t}^{(2)} = 0,$$

$$\dot{A}_{2t}^{(2)} + \Big(g_2 - \frac{\alpha}{\gamma\tau} s_{t-\tau} \Big) A_{12t}^{(2)} + (g_1 g_3 + g_2 A_{11t+\tau}^{(1)}) f_t^{(1)} A_{22t}^{(2)} + g_2 f_t^{(1)} A_{23t}^{(2)} - g_2 f_t^{(1)}$$
$$+ \sigma^2 A_{12t}^{(2)} A_{1t}^{(2)} - \sigma^2 (f_t^{(1)})^2 A_{3t}^{(2)} = 0,$$

$$\dot{A}_{3t}^{(2)} + \Big(g_2 - \frac{\alpha}{\gamma\tau} s_{t-\tau} \Big) A_{13t}^{(2)} + (g_1 g_3 + g_2 A_{11t+\tau}^{(1)}) f_t^{(1)} A_{23t}^{(2)} + g_2 f_t^{(1)} A_{33t}^{(2)} + \sigma^2 A_{13t}^{(2)} A_{1t}^{(2)} = 0,$$

$$\dot{A}_{4t}^{(2)} + 2g_4 + \frac{\tau g_1}{2\alpha} \Big(g_3 - \frac{\alpha}{\tau} s_{t-\tau} \Big)^2 + \frac{\sigma^2}{2} A_{11t}^{(2)} + \frac{\tau g_1 g_3^2}{2\alpha} + \frac{\sigma^2}{2} A_{11t+\tau}^{(1)} + \Big(g_2 - \frac{\alpha}{\gamma\tau} s_{t-\tau} \Big) A_{1t}^{(2)}$$
$$+ \frac{\sigma^2}{2} (A_{1t}^{(2)})^2 + (g_1 g_3 + g_2 A_{11t+\tau}^{(1)}) f_t^{(1)} A_{2t}^{(2)} + g_2 f_t^{(1)} A_{3t}^{(2)} = 0,$$

with the terminal conditions $A_{11T-\tau}^{(2)} = A_{11T-\tau}^{(1)}$, $A_{12T-\tau}^{(2)} = 1$, $A_{22T-\tau}^{(2)} = 2g_0$, $A_{23T-\tau}^{(2)} = 1$, $A_{13T-\tau}^{(2)} = A_{33T-\tau}^{(2)} = 0$, and $A_{iT-\tau}^{(2)} = 0$ for $i = 1, 2, 3, 4$. Therefore, the $A_{ijt}^{(2)}$s are governed by matrix Riccati differential equations and are independent of the path of s, the $A_{it}^{(2)}$ for $i = 1, 2, 3$ are governed by linear ODEs and depend on price paths, and $A_{4t}^{(2)}$ depends quadratically on $s_{t-\tau}$ and $A_{1t}^{(2)}$.

The optimal portfolio weight is given by $\phi_t^* = \phi_t^M + \phi_t^{sH} + \phi_t^{pH}$, where ϕ_t^M is given by (III.25), and the hedging demands are given by

$$\phi_t^{sH} = A_{11t}^{(2)} s_t, \quad \text{and} \quad \phi_t^{pH} = A_{12t}^{(2)} \tilde{B}_{2t}^{(2)} + A_{13t}^{(2)} \tilde{B}_{3t}^{(2)} + A_{1t}^{(2)}. \tag{III.33}$$

Based on the results regarding $\tau < T - t \leq 2\tau$, we can solve the optimal control problem for $2\tau < T - t \leq 3\tau$ by following the above procedure, that is, by first developing new state variables, then writing down the HJB equation with respect to these state variables, and then solving the HJB up to the solutions to ODEs. Eventually, we can derive the optimal portfolio weights for any horizon T recursively by using forward induction steps of length τ. To save space, we omit the expressions of $\mathbb{F}^{(n)}$ and $\mathbb{G}^{(n)}$ in Proposition 6.3, which, while more technically involved for large n (the number of state variables increases geometrically with n), can be derived up to the ODEs. We refer readers to Li and Liu (2018) for the details regarding these solutions, as they solve the optimal control problem with a more general model of time delays.

In the last two equations of (6.7), $A_{i1}^{(n)}$ is deterministic and does not depend on price paths, and $\tilde{B}_i^{(n)}$ for $i = 2, \cdots, 2^{n-1}$ are linear functionals of price paths. Therefore, both ϕ^{sH} and ϕ^{pH} can be rewritten as a weighted average of the historical returns over the look-back period by following the procedure from (A.11) to (A.14). In other words, hedging demand is a weighted average of the historical returns over the look-back period.

III.5. **Stationary Condition for the Return Process of the Momentum Asset.**

Proposition III.1. *The return process of* (2.1) *and* (2.2) *is stationary if and only if* $-1 < \alpha < 1$.

Proof. The discretization of model (2.1) and (2.2) is given by

$$r_t = \alpha \frac{r_{t-1} + \cdots + r_{t-N}}{N} + (1-\alpha)(\mu + r)\Delta t + \sigma \epsilon_t, \tag{III.34}$$

where $N = \tau/\Delta t$ is a positive integer and $\epsilon_t \sim \mathcal{N}(0, \Delta t)$. Thus, the return process of the momentum asset follows a restricted AR(N) process with the same coefficient on its lagged returns. The stationary condition is determined by the characteristic equation for (III.34), which is given by

$$\Phi(X) = X^N - \frac{\alpha}{N}(X^{N-1} + X^{N-2} + \cdots + 1) = 0. \tag{III.35}$$

It is easy to check that the process of momentum m_t has the same characteristic equation (III.35), and hence that it has the same stationary condition as return's.

The return process (III.34) is stationary if and only if all the roots of (III.35) lie inside the unit circle, which is, according to Jury's test, equivalent to

$$(\mathbf{C}_1) \quad \Phi(1) > 0; \qquad (\mathbf{C}_2) \quad (-1)^N \Phi(-1) > 0;$$

additionally,

($\mathbf{C}_3$) the $(N-1)\times(N-1)$ matrices

$$A^{\pm}_{N-1}=\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0\\ -\alpha/N & 1 & \ddots & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & 1 & 0\\ -\alpha/N & \dots & & -\alpha/N & 1\end{pmatrix}\pm\begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 & -\alpha/N\\ 0 & 0 & \cdots & & -\alpha/N & -\alpha/N\\ \vdots & \vdots & & & \vdots & \vdots\\ 0 & -\alpha/N & \cdots & \cdots & -\alpha/N & -\alpha/N\\ -\alpha/N & -\alpha/N & \cdots & \cdots & \cdots & -\alpha/N\end{pmatrix}$$

are positive innerwise.

Condition ($\mathbf{C}_1$) is equivalent to $\alpha<1$. Because

$$(-1)^N\Phi(-1)=\begin{cases}1 & \text{if } N \text{ is even;}\\ 1+\frac{\alpha}{N} & \text{if } N \text{ is odd,}\end{cases}$$

condition ($\mathbf{C}_2$) is equivalent to $\alpha>-N$. It can be verified that

$$A^+_M=\begin{cases}\left(1+\frac{\alpha}{M}\right)^{\frac{M}{2}}(1-\alpha) & \text{if } M \text{ is even;}\\ \left(1+\frac{\alpha}{M}\right)^{\frac{M-1}{2}}(1-\alpha) & \text{if } M \text{ is odd,}\end{cases}$$

and

$$A^-_M=\begin{cases}\left(1+\frac{\alpha}{M}\right)^{\frac{M}{2}} & \text{if } M \text{ is even;}\\ \left(1+\frac{\alpha}{M}\right)^{\frac{M+1}{2}} & \text{if } M \text{ is odd.}\end{cases}$$

Condition ($\mathbf{C}_3$) implies that $A^+_M>0$ and $A^-_M>0$ for $M=1,\cdots,N-1$, which is equivalent to $-1<\alpha<1$.

Thus, Conditions ($\mathbf{C}_1$)-($\mathbf{C}_3$) are equivalent to $-1<\alpha<1$, which is the necessary and sufficient condition for both return and momentum being stationary. □

III.6. **Horizon Bumps.** The following corollary provides insight into horizon bumps.

Corollary III.2. *The portfolio weight ϕ^*_0 is a non-diffusion process with respect to the investment horizon T:*

$$\frac{\partial\phi^*_0}{\partial T}=(1-\gamma)\frac{\alpha}{\gamma\sigma^2}\int_{-\tau}^{0}\varphi_v\Big(\frac{dS_v}{S_v}-rdv\Big)+C_4,$$
$$=\frac{\alpha}{\gamma\sigma^2}\int_{-\tau}^{-\tau+T}\frac{\partial\hat{\omega}(u,T)}{\partial T}(\ln S_0-\ln S_u)du+\frac{\alpha}{\gamma\sigma^2}\hat{\omega}(-\tau+T,T)(\ln S_0-\ln S_{-\tau+T})+C_5,$$

where

$$\varphi_v = \begin{cases} \int_{-\tau}^{v} \frac{\partial \hat{\omega}(u,T)}{\partial T} du, & v \in [-\tau, -\tau+T], \\ \hat{\omega}(-\tau+T,T) + \int_{-\tau}^{-\tau+T} \frac{\partial \hat{\omega}(u,T)}{\partial T} du, & v \in [-\tau+T, 0], \end{cases}$$

and C_4 *and* C_5 *are constants.*

To the leading order of $1/\gamma$,

$$\frac{\partial \phi_0^*}{\partial T} = -\frac{\alpha^2}{\gamma\sigma^2\tau^2} \int_{-\tau+T}^{0} \Big(\frac{dS_v}{S_v} - rdv \Big) - \frac{\alpha}{\gamma\sigma^2\tau} \big[(1-\alpha)\mu - \alpha r \big] = -\frac{\alpha^2}{\gamma\sigma^2\tau^2} (\ln S_0 - \ln S_{-\tau+T} + C_6). \tag{III.36}$$

Proof.

$$\frac{\partial \phi_0^*}{\partial T} = (1-\gamma)\frac{\alpha}{\gamma\sigma^2} \frac{\partial \big[\int_{-\tau}^{0} \omega(v,T) \big(\frac{dS_v}{S_v} - rdv \big) \big]}{\partial T} + C_4,$$

where

$$\int_{-\tau}^{0} \omega(v,T) \Big(\frac{dS_v}{S_v} - rdv \Big) = \int_{-\tau}^{-\tau+T} \Big[\int_{u}^{0} \hat{\omega}(u,T) \Big(\frac{dS_v}{S_v} - rdv \Big) \Big] du,$$

and $C_4 = \frac{\partial C_2(T)}{\partial T}$. Thus,

$$\frac{\partial \big[\int_{-\tau}^{0} \omega(v,T) \big(\frac{dS_v}{S_v} - rdv \big) \big]}{\partial T} = \int_{-\tau}^{0} \varphi_v \Big(\frac{dS_v}{S_v} - rdv \Big),$$

where

$$\varphi_v = \begin{cases} \int_{-\tau}^{v} \frac{\partial \hat{\omega}(u,T)}{\partial T} du, & v \in [-\tau, -\tau+T], \\ \hat{\omega}(-\tau+T,T) + \int_{-\tau}^{-\tau+T} \frac{\partial \hat{\omega}(u,T)}{\partial T} du, & v \in [-\tau+T, 0]. \end{cases}$$

In addition, $\frac{\partial \phi^*}{\partial T}$ can be written in terms of cumulative returns as follows:

$$\frac{\partial \phi_0^*}{\partial T} = \int_{-\tau}^{-\tau+T} \frac{\alpha}{\gamma\sigma^2} \frac{\partial \hat{\omega}(u,T)}{\partial T} (\ln S_0 - \ln S_u) du + \frac{\alpha}{\gamma\sigma^2} \hat{\omega}(-\tau+T,T)(\ln S_0 - \ln S_{-\tau+T}) + C_5,$$

where $C_5 = \frac{\alpha(\tau-T)\hat{\omega}(-\tau+T,T)}{2\gamma} - \frac{\alpha}{2\gamma} \int_{-\tau}^{-\tau+T} \frac{\partial \hat{\omega}(u,T)}{\partial T} u du$. To the leading order of $1/\gamma$, $\frac{\partial \phi_0^*}{\partial T}$ is given by (III.36), where

$$C_6 = -\sigma^2(T-\tau)/2 + [(1-\alpha)\mu - \alpha r]\tau/\alpha. \tag{III.37}$$

□

In (III.36), the sign of the cumulative return over $[-\tau+T, 0]$ changes as T decreases to zero, leading to bumps in the horizon dependence. (Although $\frac{dS_v}{S_v}$ changes signs an infinite number of times within any finite interval, $\frac{\partial \phi^*}{\partial T}$ may change signs less frequently depending on the recent price trend over $[-\tau+T, 0]$.)

III.7. **Optimal Portfolio Weights for Large γ.** To provide more insight into optimal portfolio weights, we study their approximation to the leading order of $1/\gamma$. This exercise leads to a more explicit expression of ω_v.

Corollary III.3. *The weight ω_v on the historical instantaneous excess return $dS_v/S_v - rdv$ in ϕ_0^H is (to the leading order of $1/\gamma$) given by*

$$\omega_v = \begin{cases} (\tau+v)/\tau^2, & v \in [-\tau, -\tau+T], \\ T/\tau^2, & v \in [-\tau+T, 0], \end{cases}$$

and ϕ_0^H is (to the leading order of $1/\gamma$) given by

$$\phi_0^H = -\frac{\alpha^2}{\gamma\sigma^2}\Big\{\int_{-\tau}^{-\tau+T} \frac{v+\tau}{\tau^2}\Big(\frac{dS_v}{S_v} - rdv\Big) + \int_{-\tau+T}^{0} \frac{T}{\tau^2}\Big(\frac{dS_v}{S_v} - rdv\Big) + \frac{1-\alpha}{\alpha\tau}\mu T - \frac{rT^2}{2\tau^2}\Big\}. \tag{III.38}$$

Proof. When $\gamma \to \infty$, to the leading order of $1/\gamma$, we have $\hat{\omega} = \alpha\gamma^{-1}\tau^{-2}$. The weight ω_v on the historical instantaneous return dS_v/S_v becomes

$$\omega_v = \begin{cases} \frac{\alpha}{\gamma\tau^2}(v+\tau), & v \in [-\tau, -\tau+T], \\ \frac{\alpha}{\gamma\tau^2}T, & v \in [-\tau+T, 0], \end{cases} \tag{III.39}$$

and ϕ^H reduces to (III.38). □

Corollary III.3 shows that the weight ω_v on the historical excess return $dS_v/S_v - rdv$ is linearly increasing in v for $v \in [-\tau, -\tau+T]$. In the general case, Figure 2 shows that the weight ω_v in (3.8) is an approximately linear function of v for $v \in [-\tau, -\tau+T]$, suggesting that (III.38) in Corollary III.3 provides a close approximation of ϕ_0^H.

The cumulative return of the momentum asset depends more heavily on its recent returns than its distant ones, as shown in Online Appendix I. Equation (III.38) shows that ϕ_0^H depends even more heavily on recent versus distant returns.

Corollary III.4. *The optimal portfolio weight of the momentum asset can be (to the leading order of $1/\gamma$) rewritten as*

$$\phi_0^* = \frac{\alpha m_0^M + (1-\alpha)\mu}{\gamma\sigma^2}, \tag{III.40}$$

where

$$m_0^M = \frac{1}{\tau}\Big[\int_{-\tau}^{-\tau+T}\Big(1-\alpha-\alpha\frac{v}{\tau}\Big)\Big(\frac{dS_v}{S_v} - rdv\Big) + \int_{-\tau+T}^{0}\Big(1-\alpha\frac{T}{\tau}\Big)\Big(\frac{dS_v}{S_v} - rdv\Big) - (1-\alpha)\mu T + \frac{\alpha r T^2}{2\tau}\Big].$$

Practitioners often use a mean-variance portfolio. Corollary III.4 shows that the optimal momentum strategy can also be treated as a mean-variance strategy that uses a horizon- and path-adjusted momentum variable m_0^M to predict future returns. Different from m_0, which is an equally weighted MA of past returns, m_0^M places less weight on recent returns. In addition, the weights of past returns depend on the length of horizon T. For short T, m_0^M is approximately equal to m_0. For long T, after adjusting for the horizon and price path, m_0^M can be significantly different from m_0. For example, they may have different signs after a rebound price path than they do after another type of path.

The difference between m_0^M and m_0 reflects the path dependence of the tradeoff between good and bad states. An investor who is more risk averse than a log-utility investor values the stock relatively more in states of the world in which expected returns are low because her marginal utility is higher. Accordingly, such an investor holds an adjusted mean-variance portfolio by valuing the stock more in "bad" states; these good and bad states are determined by the paths in (III.40), different from the standard mean-variance portfolio.

III.8. **Optimal Portfolio Weight in Terms of Cumulative Returns.**

Corollary III.5. *The optimal portfolio weight of* (3.8) *can also be written as*

$$\phi_0^* = \frac{\alpha}{\gamma\sigma^2}\left[\frac{\ln S_0 - \ln S_{-\tau}}{\tau} + \int_{-\tau}^{-\tau+T}(1-\gamma)\hat{\omega}_u(\ln S_0 - \ln S_u)du\right] + C_2, \tag{III.41}$$

where $\hat{\omega}_u > 0$ *is a deterministic weight given by* (A.14) *and* C_2 *is a constant given by* (III.43).

Thus, the optimal portfolio weight is (to the leading order of $1/\gamma$*) given by*

$$\begin{aligned}\phi_0^* =& \frac{\alpha}{\gamma\sigma^2}\left[\frac{\ln S_0 - \ln S_{-\tau}}{\tau} - \int_{-\tau}^{-\tau+T}\frac{\alpha}{\tau^2}(\ln S_0 - \ln S_u)du\right] \\ &+ \frac{(\tau-\alpha T)[(1-\alpha)\mu - \alpha r]}{\gamma\sigma^2\tau} + \frac{\alpha}{2\gamma} + \frac{\alpha^2 T(T-2\tau)}{4\gamma\tau^2}.\end{aligned} \tag{III.42}$$

Proof. Using (3.8),

$$\begin{aligned}\phi_0^* &= \phi_0^M + (1-\gamma)\frac{\alpha}{\gamma\sigma^2}\int_{-\tau}^{0}\omega_v\frac{dS_v}{S_v} + C_1 = \phi_0^M + (1-\gamma)\frac{\alpha}{\gamma\sigma^2}\int_{-\tau}^{-\tau+T}\int_u^0\frac{dS_v}{S_v}\hat{\omega}_u du + C_1 \\ &= \frac{\alpha}{\gamma\sigma^2\tau}(\ln S_0 - \ln S_{-\tau}) + (1-\gamma)\frac{\alpha}{\gamma\sigma^2}\Big(\int_{-\tau}^{-\tau+T}\hat{\omega}_u du \ln S_0 - \int_{-\tau}^{-\tau+T}\hat{\omega}_u \ln S_u du\Big) + C_2,\end{aligned}$$

where C_2 is a constant given by

$$C_2 = C_1 - (1-\gamma)\frac{\alpha}{\gamma\sigma^2}\int_{-\tau}^{-\tau+T}\frac{\sigma^2 u\hat{\omega}_u}{2}du + \frac{(1-\alpha)\mu - \alpha r}{\gamma\sigma^2} + \frac{\alpha}{2\gamma}. \tag{III.43}$$

□

Corollary III.5 states that the optimal portfolio weight is a weighted average of historical cumulative returns with a start time that varies from $-\tau$ to $-\tau + T$ and with an end time of 0. In (III.41) and (III.42), the first component in the square bracket corresponds to the mean-variance portfolio weight, which is dependent upon the cumulative return over the look-back period. The second component (i.e., hedging) is a weighted average of the historical cumulative returns over different historical periods. For large γ, (III.42) shows that the second component depends on an equally weighted average of the historical cumulative returns.

III.9. **Price Level Independence.**

Corollary III.6. *When the historical price path s_u is changed to $s_u + c$ for all $u \in [-\tau, 0]$ (where c is a constant), ϕ^M and ϕ^H do not change. Thus, both of the demand components depend on historical returns.*

Proof. Changes in the price level do not affect the price trend $s_0 - s_{-\tau}$; thus, according to (I.2), the mean-variance portfolio weight does not change.

It follows from (3.3), (3.4) and (A.7) that constant change in a path does not affect $\mathbb{E}_t[\xi_T^{(\gamma-1)/\gamma}]$. Moreover, (A.4) implies that this change affects W_t only via W_0. Therefore, this change does not affect the optimal portfolio weight.

Thus, both the mean-variance weight and the optimal portfolio weight are not affected by a change in the level of historical prices. Because the weight of historical price s_u in the optimal portfolio is different for different u, the optimal portfolio is affected by historical returns. □

III.10. **Monte Carlo Simulation Method for $T - t > \tau$.**

Proposition 6.3 shows that the optimal portfolio weight for longer horizons can be recursively solved up to the solutions to ODEs. In this section, we also provide a numerical method of solving for the optimal portfolio weight.

The conditional expectations are calculated using the least squares Monte Carlo approach (Longstaff and Schwartz, 2001). More specifically, we simulate 10,000 time series of price date points over $[t, T]$ for a given historical path during $[t - \tau, t]$, which are generated by model (2.1) and (2.2). The conditional expectation $E_t\big[\xi_T^{(\gamma-1)/\gamma}\big]$ in Proposition 3.1 is the average of $\xi_T^{(\gamma-1)/\gamma}$. The conditional expectation $E_{t+dt}\big[\xi_T^{(\gamma-1)/\gamma}\big]$ is derived by regressing the realizations of $\xi_T^{(\gamma-1)/\gamma}$ on a constant and on the corresponding shocks $d\hat{B}_t$ at time $t + dt$, following Longstaff and Schwartz (2001). We find that adding more regressors, such as $(d\hat{B}_t)^2$, $(d\hat{B}_t)^3$, or prices (such as $\hat{s}_{t+dt}$, $\hat{s}^2_{t+dt}$ or $\hat{s}^3_{t+dt}$), does not affect the results. Then, ψ_t in (3.7) can be derived by regressing $d(\pi_t W_t) = W_0 \bar{\xi}_0^{-1}(\mathbb{E}_{t+dt}[\xi_T^{(\gamma-1)/\gamma}] - \mathbb{E}_t[\xi_T^{(\gamma-1)/\gamma}])$ on $d\hat{B}_t$. The optimal portfolio weights follow from (3.6).

Appendix IV. Empirical Analysis

To assess the empirical relevance of our theoretical results, we calibrate our model to U.S. factors and examine the performance of the optimal momentum strategy. We show that the optimal strategy generates returns with higher means, lower volatility, and hence much higher Sharpe ratios than those generated by the mean-variance strategy widely used in the literature that exploits momentum alone and ignores paths. By exploiting the paths of the factors, the optimal strategy also significantly increases return skewness and decreases return kurtosis. We find that the outperformance of the optimal strategy tends to be the highest after extreme periods with large price swings, highlighting the effect of path dependence.

IV.1. **Model Estimation.** To ensure consistency with the momentum literature, we discretize the continuous-time model (2.1) at a monthly frequency:

$$R_{t+1} = (1-\alpha)\mu + \alpha m_t + \sigma\epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0,1), \tag{IV.1}$$

where R_{t+1} is the momentum asset's return in excess of the short rate.

Ehsani and Linnainmaa (2021) found evidence that factor returns exhibit time series momentum and that time series factor momentum fully subsumes cross-sectional momentum in the context of individual stock returns. Following Ehsani and Linnainmaa (2021), we calibrate model (IV.1) according to common U.S. factors. The five factors of Fama and French (2015), as well as the short-term reversal and long-term reversal factors, are downloaded from Kenneth French's website at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/`, and the betting-against-beta and quality-minus-junk factors are downloaded from the AQR data sets at `https://www.aqr.com/insights/datasets`. For the factors of which the data are not provided, we compute the factor returns as the average return of the three top deciles minus that of the three bottom deciles, and the portfolio data for this calculation are from Kenneth French's website. The examined anomalies include earnings to price, cash flow to price, dividend yield, accruals, market beta, net share issues, variance, and residual variance. To ensure consistency with the optimal dynamic strategies that require nonzero investment (wealth), for each factor, we place investments in both long and short legs in amounts equal to that of the initial wealth at the beginning of the time period and that of the remaining wealth invested in riskless assets. We view this portfolio as our momentum asset. In addition to considering the common risk factors, we calibrate our model to U.S. stocks from different industries.

We set the short rate as $r = 0.04$ annually. We also set $\tau = 1$; i.e., m_t is the MA of the historical excess returns over the previous year. We calibrate model (IV.1) using the maximum likelihood method. Table IV.1 reports the estimated parameters.

Almost all the estimates of α are significantly positive, showing that the returns of these factors exhibit time series momentum. This confirms the results of Ehsani and Linnainmaa (2021), who find that time series momentum (defined as the MA of historical returns) is a pervasive feature of factor returns.

IV.2. **Performance Analysis.** In this subsection, we assess the performance of the optimal momentum strategy based on the given factors using the estimated parameters in Table IV.1. We set $\gamma = 2$ and $T = 3$. We construct the optimal momentum portfolios for each month and hold them for 3 years. Rebalancing is assumed to take place monthly.

For all the factors, the optimal strategy invests less aggressively with lower portfolio weight in the risky asset than the mean-variance strategy does, as shown in Column 3 of Table IV.2. Column 5 reports the t-statistic of the coefficient from a regression of $\Delta\phi = |\phi^* - \phi^M|$ on X_{max}. Confirming our predictions, the difference between the optimal portfolio weight and the mean-variance weight tends to be the largest during the extreme periods with large price swings.

Table IV.2 also reports the portfolio performance. For all the assets, the optimal momentum strategy generates returns with higher means, lower volatility, and hence much higher Sharpe ratios than the mean-variance strategy that exploits momentum alone and ignores paths. For example, when investing in the size factor, the optimal momentum strategy generates annualized returns with a mean of 5.64% and a standard deviation of 4.31%, producing a Sharpe ratio of 0.38. Comparatively, the mean-variance strategy generates annualized returns with a mean of 5.60% and a standard deviation of 13.26%, producing a Sharpe ratio of only 0.12. This is consistent with our theory, which shows that optimal re-balancing rules also implicitly maximize Sharpe ratios (see (A.6) in Appendix A.2).

Mean-variance momentum strategies occasionally crash; and this is associated with negatively skewed return distributions. For example, Daniel and Moskowitz (2016) document stock momentum crashes, and Ehsani and Linnainmaa (2021) show that crashes occur when factor returns abruptly become negatively autocorrelated. By exploiting path dependence, our optimal dynamic momentum strategy positively skews all the factors and stocks. In addition, in most cases, the return skewness of the optimal strategy is much higher than that of both the mean-variance strategy and the factor. For example, return skewness of the size factor is 0.40. Applying the optimal strategy and the mean-variance strategy to the size factor produces return skewness of 1.43 and -0.68, respectively. The optimal strategy also generates much lower return kurtosis than the mean-variance strategy. The return kurtosis of the size factor, the optimal strategy, and the mean-variance strategy are 6.13, 4.69, and 21.33, respectively. Thus, the optimal strategy tends to produce infrequent large profits instead of crashes, suggesting that the optimal strategy helps avoid momentum crashes. Indeed, an investor with a risk aversion coefficient greater than one values risky assets relatively more in states of the world where expected returns are lower because her marginal utility is higher. As a result, optimal strategies tend to increase return skewness by preventing large losses.

To examine path dependence, we regress portfolio return onto X_{max}. Table IV.2 shows that the return of the optimal strategy is significantly and positively related to X_{max}, but the mean-variance strategy return and the asset return are not related to this variable. Because portfolio weights depend on past returns, one may expect that the strategy itself could produce path-dependent performance. To address this concern, we apply the momentum strategy to a simulated IID return series with a mean and a standard deviation equal to those of the assets. Table IV.2 shows that the portfolio returns are not significantly related to X_{max}. Therefore, the observed path dependence is due to momentum.

To focus on momentum, we do not account for transaction costs in our optimal strategy. Optimal portfolio selection with transaction costs has been studied by, e.g., Liu and Loewenstein (2002), Dai, Zhang and Zhu (2010), and Dai, Yang, Zhang and Zhu (2016). In particular, Dai, Zhang and Zhu (2010) and Dai et al. (2016) study the optimal trend-following strategy with transaction costs where market trends are characterized with a Markov regime-switching model and show that the optimal strategy can be profitable after accounting for reasonable transaction costs. Instead of theoretically studying the interaction between momentum and transaction costs, we numerically examine the performance of the optimal momentum strategy assuming a proportional transaction cost of 0.5%. Table IV.2 shows that the optimal strategy can still be profitable in this case.

Table IV.1. Parameter estimations

This table reports the sample periods of 17 U.S. factors as well as those of some U.S. stocks, the estimated parameters of the momentum model (2.1) and (2.2) in annual terms, and the related t-statistics (in brackets) based on the factor returns using the maximum likelihood method.

		Estimations		
Factor	Sample period	α	μ	σ
Market	1963:07–2019:12	0.14	0.08	0.64
		[1.41]	[3.52]	[47.10]
Size	1963:07–2019:12	0.34	0.07	0.36
		[2.85]	[3.53]	[36.50]
Value	1963:07–2019:12	0.34	0.08	0.34
		[3.34]	[3.96]	[36.50]
Profitability	1963:07–2019:12	0.34	0.08	0.26
		[3.25]	[5.12]	[36.50]
Investment	1963:07–2019:12	0.39	0.08	0.24
		[3.83]	[5.12]	[36.50]
Short-term reversal	1926:07–2019:12	0.30	0.12	0.41
		[3.27]	[6.96]	[47.10]
Long-term reversal	1931:01–2019:12	0.42	0.07	0.41
		[4.91]	[3.05]	[45.96]
Betting against beta	1930:12–2019:12	0.49	0.12	0.38
		[6.65]	[5.31]	[45.98]
Quality minus junk	1957:07–2019:12	0.37	0.09	0.26
		[3.69]	[5.84]	[38.42]
Earnings to price	1951:07–2019:12	0.29	0.09	0.33
		[3.02]	[5.33]	[40.25]
Cashflow to price	1951:07–2019:12	0.26	0.08	0.33
		[2.55]	[4.88]	[40.25]
Dividend yield	1927:07–2019:12	0.21	0.05	0.44
		[2.21]	[2.86]	[46.86]
Accruals	1963:07–2019:12	0.15	0.07	0.23
		[1.24]	[6.80]	[36.50]
Market beta	1963:07–2019:12	0.09	0.06	0.56
		[0.73]	[2.30]	[36.50]
Net share issues	1963:07–2019:12	0.42	0.07	0.29
		[4.15]	[3.76]	[36.50]
Variance	1963:07–2019:12	0.22	0.06	0.65
		[1.87]	[1.83]	[36.50]
Residual variance	1963:07–2019:12	0.25	0.06	0.60
		[2.19]	[1.92]	[36.50]
Stock	Sample period	α	μ	σ
Amazon	1997:06–2019:12	0.46	0.30	1.97
		[2.79]	[1.26]	[22.75]
Apple	1981:01–2019:12	0.09	0.29	1.55
		[0.59]	[3.56]	[30.20]
Coca-Cola	1927:07–2019:12	0.22	0.14	0.72
		[2.26]	[5.01]	[46.86]
General Motors	1927:07–2019:12	0.37	0.08	1.08
		[4.34]	[1.58]	[46.48]

Table IV.2. Portfolio performance

This table reports the means and standard deviations (SDs) of the portfolio weights, and the means (in percentages), standard deviations (in percentages), skewness (Skew), kurtosis (Kurt), and Sharpe ratios (SRs) of the returns for the following portfolios: the optimal portfolio, the mean-variance portfolio, the momentum asset (17 U.S. factors and 4 U.S. stocks), the optimal portfolio accounting for a proportional transaction cost of 0.5%, and the portfolio resulting from applying the momentum strategy to IID returns. We report the t-statistics for coefficients $b^{\Delta\phi}$ and b^R in parenthesis from regressions $\Delta\phi_t = a + b^{\Delta\phi} X_{max,t} + \varepsilon_t$ and $R_{t+1} = a + b^R X_{max,t} + \varepsilon_{t+1}$, respectively, where $\Delta\phi = |\phi^* - \phi^M|$ is the deviation of the optimal portfolio weight from the mean-variance weight, R is the portfolio return, and X_{max} captures extreme periods as defined in (4.1). The standard errors are adjusted for heteroskedasticity and serial correlation according to Newey and West (1987). All the results are in annual terms.

Factor	Strategy	Portfolio weight			Portfolio return					
		Mean	SD	$t(b^{\Delta\phi})$	Mean	SD	Skew	Kurt	SR	$t(b^R)$
Market	Optimal momentum strategy	0.09	0.03	(2.22)	4.71	3.14	0.37	4.64	0.23	(1.80)
	Mean-variance strategy	0.10	0.03		4.45	5.98	0.07	13.85	0.08	(-2.55)
	Factor return				7.92	64.14	0.18	10.77	0.06	(0.95)
	Transaction costs				4.09	2.85	0.42	4.71	0.03	(1.73)
	IID return				4.09	6.18	0.03	4.11	0.01	(1.11)
Size	Optimal momentum strategy	0.27	0.18	(5.68)	6.09	11.86	1.15	3.23	0.18	(4.82)
	Mean-variance strategy	0.29	0.15		5.60	13.26	-0.68	21.33	0.12	(1.14)
	Factor return				7.46	36.41	0.40	6.13	0.10	(1.72)
	Transaction costs				4.14	9.53	1.15	3.38	0.01	(4.36)
	IID return				4.40	10.63	-0.10	4.91	0.04	(-1.62)
Value	Optimal momentum strategy	0.32	0.22	(3.31)	6.93	11.70	1.14	4.39	0.25	(2.27)
	Mean-variance strategy	0.35	0.19		6.26	15.01	0.72	17.22	0.15	(-1.75)
	Factor return				8.08	34.32	0.15	4.91	0.12	(0.26)
	Transaction costs				4.53	9.34	1.19	4.65	0.06	(1.96)
	IID return				4.69	11.95	-0.06	4.60	0.06	(-0.02)
Profitability	Optimal momentum strategy	0.54	0.32	(1.03)	7.78	14.99	1.10	5.03	0.25	(2.70)
	Mean-variance strategy	0.57	0.24		6.95	17.38	1.23	17.92	0.17	(0.25)
	Factor return				7.74	26.21	-0.31	15.10	0.14	(0.99)
	Transaction costs				3.84	11.29	1.04	5.14	-0.01	(2.31)
	IID return				5.31	15.04	-0.01	3.97	0.09	(-0.69)
Investment	Optimal momentum strategy	0.63	0.45	(6.90)	8.97	22.42	1.41	4.75	0.22	(2.56)
	Mean-variance strategy	0.67	0.31		7.69	21.82	1.01	14.61	0.17	(-1.32)
	Factor return				7.85	24.44	0.38	4.45	0.16	(-0.41)
	Transaction costs				4.10	16.52	1.39	4.78	0.01	(1.64)
	IID return				5.60	16.29	0.00	4.08	0.10	(0.11)

Table IV.2. continued

Factor	Strategy	Portfolio weight			Portfolio return					
		Mean	SD	$t(b^{\Delta\phi})$	Mean	SD	Skew	Kurt	SR	$t(b^R)$
Short-term reversal	Optimal momentum strategy	0.33	0.13	(8.89)	8.32	19.35	3.02	14.59	0.22	(2.12)
	Mean-variance strategy	0.36	0.12		7.41	18.46	1.96	35.92	0.18	(1.62)
	Factor return				12.05	41.00	0.95	11.11	0.20	(0.59)
	Transaction costs				5.81	16.44	3.05	14.72	0.11	(1.77)
	IID return				6.20	14.93	0.09	3.89	0.15	(0.12)
Long-term reversal	Optimal momentum strategy	0.19	0.22	(3.46)	5.77	12.11	6.06	61.62	0.15	(0.86)
	Mean-variance strategy	0.20	0.19		5.68	20.50	9.03	149.36	0.08	(-0.76)
	Factor return				6.61	41.39	2.87	27.57	0.06	(-0.95)
	Transaction costs				4.26	10.15	6.40	69.16	0.03	(0.29)
	IID return				4.06	9.50	-0.04	7.52	0.01	(1.51)
Betting against beta	Optimal momentum strategy	0.38	0.33	(2.05)	10.85	28.63	2.27	9.18	0.24	(5.40)
	Mean-variance strategy	0.42	0.27		9.40	21.24	0.04	15.48	0.25	(1.43)
	Factor return				11.94	38.70	-0.71	10.55	0.21	(1.32)
	Transaction costs				7.54	23.31	2.30	9.40	0.15	(4.74)
	IID return				6.43	16.51	0.14	5.15	0.15	(0.67)
Quality minus junk	Optimal momentum strategy	0.61	0.35	(5.81)	9.06	18.00	0.58	3.23	0.28	(-0.65)
	Mean-variance strategy	0.65	0.26		8.09	21.38	0.31	10.96	0.19	(0.04)
	Factor return				8.84	26.35	0.17	6.02	0.18	(-0.45)
	Transaction costs				4.53	13.71	0.47	3.42	0.04	(-0.93)
	IID return				6.15	16.79	0.02	3.81	0.13	(1.27)
Earnings to price	Optimal momentum strategy	0.38	0.18	(5.75)	7.13	12.25	0.82	3.10	0.26	(2.11)
	Mean-variance strategy	0.41	0.16		6.44	14.92	0.31	11.18	0.16	(1.03)
	Factor return				8.63	32.69	0.19	5.55	0.14	(1.50)
	Transaction costs				4.37	9.91	0.86	3.19	0.04	(1.63)
	IID return				5.03	13.11	-0.02	3.82	0.08	(-0.39)
Cashflow to price	Optimal momentum strategy	0.34	0.15	(7.52)	6.26	9.85	0.86	3.47	0.23	(1.36)
	Mean-variance strategy	0.36	0.14		5.68	13.00	-0.54	16.11	0.13	(1.19)
	Factor return				7.60	32.71	0.04	5.46	0.11	(1.50)
	Transaction costs				3.87	7.94	0.89	3.57	-0.02	(0.99)
	IID return				4.54	11.47	-0.04	3.85	0.05	(-1.48)

Table IV.2. continued

Factor	Strategy	Portfolio weight			Portfolio return					
		Mean	SD	$t(b^{\Delta\phi})$	Mean	SD	Skew	Kurt	SR	$t(b^R)$
Dividend yield	Optimal momentum strategy	0.12	0.08	(4.37)	4.65	4.42	1.77	7.89	0.15	(0.83)
	Mean-variance strategy	0.12	0.08		4.33	8.56	1.35	54.80	0.04	(-2.22)
	Factor return				4.87	44.49	1.01	12.87	0.02	(1.84)
	Transaction costs				3.84	3.79	1.83	9.44	-0.04	(0.59)
	IID return				3.84	5.87	-0.08	6.05	-0.03	(0.94)
Accruals	Optimal momentum strategy	0.65	0.11	(9.13)	6.72	10.14	-0.03	2.48	0.27	(0.66)
	Mean-variance strategy	0.67	0.10		6.26	15.95	-0.37	5.31	0.14	(-1.73)
	Factor return				7.15	23.14	-0.16	4.62	0.14	(-1.80)
	Transaction costs				2.32	8.13	0.00	2.53	-0.21	(0.34)
	IID return				5.28	15.51	0.02	3.10	0.08	(-0.30)
Market beta	Optimal momentum strategy	0.09	0.02	(0.36)	4.41	2.69	0.29	3.20	0.15	(-1.40)
	Mean-variance strategy	0.09	0.03		4.17	5.01	0.33	8.72	0.03	(-0.85)
	Factor return				5.49	56.13	0.18	5.04	0.03	(-1.69)
	Transaction costs				3.84	2.46	0.29	3.33	-0.07	(-1.36)
	IID return				3.88	4.92	-0.04	3.69	-0.02	(-0.65)
Net share issues	Optimal momentum strategy	0.42	0.40	(5.46)	7.67	19.47	2.14	9.74	0.19	(4.00)
	Mean-variance strategy	0.45	0.28		6.75	20.75	3.02	31.89	0.13	(1.18)
	Factor return				7.23	28.88	0.83	6.80	0.11	-0.15
	Transaction costs				4.26	14.96	2.09	9.67	0.02	(3.48)
	IID return				4.67	13.04	-0.08	4.97	0.05	-0.22
Variance	Optimal momentum strategy	0.07	0.05	(0.11)	4.69	3.20	0.32	2.85	0.22	(-0.52)
	Mean-variance strategy	0.07	0.06		4.41	7.46	-0.98	17.75	0.05	(-2.59)
	Factor return				5.96	64.89	-0.29	5.45	0.03	(0.54)
	Transaction costs				4.20	2.80	0.33	3.55	0.07	(-1.02)
	IID return				3.90	5.10	-0.33	6.35	-0.02	(0.17)
Residual variance	Optimal momentum strategy	0.08	0.07	(0.00)	4.84	3.80	0.21	2.34	0.22	(0.27)
	Mean-variance strategy	0.08	0.07		4.54	8.35	-0.82	17.97	0.06	(-2.74)
	Factor return				6.07	60.53	-0.30	7.01	0.03	(-0.25)
	Transaction costs				4.25	3.28	0.12	2.90	0.08	(-0.07)
	IID return				3.91	5.64	-0.35	6.48	-0.02	(-0.01)

Table IV.2. continued

Stock	Strategy	Portfolio weight			Portfolio return					
		Mean	SD	$t(b^{\Delta\phi})$	Mean	SD	Skew	Kurt	SR	$t(b^{R})$
Amazon	Optimal momentum strategy	0.03	0.04	(1.62)	5.15	5.27	3.45	19.69	0.22	(0.88)
	Mean-variance strategy	0.05	0.05		7.26	21.26	5.34	50.90	0.15	(-1.46)
	Stock return				40.14	200.06	1.97	15.31	0.18	(-1.01)
	Transaction costs				4.91	5.04	3.40	19.53	0.18	(0.77)
	IID return				5.00	9.53	-0.34	6.75	0.10	(1.50)
Apple	Optimal momentum strategy	0.06	0.01	(-0.59)	5.87	5.59	0.62	2.86	0.33	(-1.64)
	Mean-variance strategy	0.06	0.01		5.46	9.47	-0.31	5.39	0.15	(-1.50)
	Stock return				28.83	155.60	-0.09	4.31	0.16	(-0.25)
	Transaction costs				5.47	5.46	0.61	2.86	0.27	(-1.65)
	IID return				4.97	9.16	-0.04	3.12	0.11	(-0.96)
Coca-Cola	Optimal momentum strategy	0.13	0.05	(-0.96)	5.99	9.51	1.98	8.81	0.21	(1.33)
	Mean-variance strategy	0.13	0.05		5.56	10.14	0.74	7.55	0.15	(0.55)
	Stock return				13.84	71.86	-0.08	5.41	0.14	(0.01)
	Transaction costs				5.11	8.73	2.01	8.99	0.13	(1.25)
	IID return				4.83	9.68	0.05	4.14	0.09	(0.34)
General Motors	Optimal momentum strategy	0.03	0.06	(3.74)	5.44	12.28	4.47	23.52	0.12	(2.67)
	Mean-variance strategy	0.04	0.06		5.04	11.61	7.84	156.12	0.09	(-0.79)
	Stock return				8.51	109.29	0.41	14.08	0.04	(-0.24)
	Transaction costs				5.13	11.64	4.51	23.85	0.10	(2.57)
	IID return				3.90	6.28	-0.17	9.92	-0.02	(-0.93)