Journal of Financial Economics 92 (2009) 519–544

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Technical analysis: An asset allocation perspective on the use of moving averages[☆]

Yingzi Zhu[a], Guofu Zhou[b,c,*]

[a] School of Economics and Management, Tsinghua University, Beijing 100084, China
[b] Olin Business School, Washington University in St. Louis, St. Louis, MO 63130, USA
[c] China Center for Financial Research, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 17 September 2007
Received in revised form
21 July 2008
Accepted 24 July 2008
Available online 9 March 2009

JEL classification:
G11
G12
G14
C11
C61

Keywords:
Technical analysis
Trading rules
Asset allocation
Predictability
Learning

ABSTRACT

In this paper, we analyze the usefulness of technical analysis, specifically the widely employed moving average trading rule from an asset allocation perspective. We show that, when stock returns are predictable, technical analysis adds value to commonly used allocation rules that invest fixed proportions of wealth in stocks. When uncertainty exists about predictability, which is likely in practice, the fixed allocation rules combined with technical analysis can outperform the prior-dependent optimal learning rule when the prior is not too informative. Moreover, the technical trading rules are robust to model specification, and they tend to substantially outperform the model-based optimal trading strategies when the model governing the stock price is uncertain.

1. Introduction

Technical analysis uses past prices and perhaps other past statistics to make investment decisions. Proponents of technical analysis believe that these data contain important information about future movements of the stock market. In practice, all major brokerage firms publish technical commentary on the market and many of the advisory services are based on technical analysis. In his interviews with them, Schwager (1993, 1995) finds

☆ This paper supersedes an earlier version circulated under the title "Technical Analysis and Theory of Finance." We are grateful to Kerry Back, William Brock, Henry Cao, Bjørn Eraker, Ramo Gencay, Ravi Jagannathan, Raymond Kan, Hong Liu, Jun Liu, Lars Lochstoer, Chris Neely, Luboš Pástor, G. William Schwert (the managing editor), Jack Strauss, Avanidhar Subrahmanyam, Tan Wang, seminar participants at Fudan University, Renmin University of China, Tulane University, Tsinghua University, and Washington University, participants at the 2007 China International Conference in Finance and the 2007 European Finance Association annual meetings, and especially an anonymous referee for insightful and detailed comments that substantially improved the paper. Guofu Zhou acknowledges China Center for Financial Research of Tsinghua University, where part of the paper was written during a May 2006 visit. Yingzi Zhu acknowledges financial support from Minoru Kobayashi China Economic Research Fund. We also thank Wei Chen, Xiaobo Liu, and Jinfan Zhang for research assistance, and Lynnea Brumbaugh-Walter for many helpful editorial comments.
* Corresponding author at: Olin Business School, Washington University in St. Louis, St. Louis, MO 63130, USA. Tel.: +1 314 935 6384; fax: +1 314 935 6359.
E-mail address: zhou@wustl.edu (G. Zhou).

0304-405X/$ - see front matter
doi:10.1016/j.jfineco.2008.07.002

that many top traders and fund managers use it. Moreover, Covel (2005), citing examples of large and successful hedge funds, advocates the use of technical analysis exclusively without learning any fundamental information on the market.

Academics, however, have long been skeptical about the usefulness of technical analysis, despite its widespread acceptance and adoption by practitioners.[1] There are perhaps three reasons. The first reason is that no theoretical basis exists for it, which this paper attempts to provide. The second reason is that earlier theoretical studies often assume a random walk model for the stock price, which completely rules out any profitability from technical trading. The third reason is that earlier empirical findings, such as Cowles (1933) and Fama and Blume (1966), are mixed and inconclusive. Recently, however, Brock, Lakonishok, and LeBaron (1992), and especially Lo, Mamaysky, and Wang (2000), find strong evidence of profitability in technical trading based on more data and more elaborate strategies. These studies stimulated much subsequent academic research on technical analysis, but these later studies focus primarily on the statistical validity of the earlier results.

Our paper takes a new perspective. We consider the theoretical rationales for using technical analysis in a standard asset allocation problem. An investor chooses how to allocate his wealth optimally between a riskless asset and a risky one, which we call stock. For tractability, we focus on the profitability of the simplest and seemingly the most popular technical trading rule (the moving average, MA), which suggests that investors buy the stock when its current price is above its average price over a given period L.[2] The immediate question is what proportion of wealth the investor should allocate into the stock when the MA signals so. Previous studies use an all-or-nothing approach: the investor invests 100% of his wealth into the stock when the MA says buy and nothing otherwise. This common and naive use of the MA is not optimal from an asset allocation perspective because the optimal amount should be a function of the investor's risk aversion as well as the degree of predictability of the stock return. Intuitively, if the investor invests an optimal fixed proportion of his money into the stock market, say 80%, when there is no MA signal, he should invest more than 80% when the MA signals a buy and less otherwise. The 100% allocation is clearly unlikely to be optimal. We solve, for a log-utility investor, the problem of allocating the optimal amount of stock explicitly, which provides a clear picture of how the degree of predictability affects the allocation decision given the log-utility risk tolerance. We also solve the optimal investment problem both approximate analytically and via simulations in the more general power-utility case. The results show that the use of the MA can help increase the investor's utility substantially.

Moreover, given an investment strategy that allocates a fixed proportion of wealth to the stock, we show that the MA rule can be used in conjunction with the fixed rule to yield higher expected utility. In particular, it can improve the expected utility substantially for the popular fixed strategy that follows the Markowitz (1952) modern portfolio theory and the Tobin (1958) two-fund separation theorem. Because indexing, a strategy of investing in a well-diversified portfolio of stocks, makes up roughly one-third of the US stock market, and its trend is on the rise worldwide (see, e.g., Bhattacharya and Galpin, 2006), and because popular portfolio optimization strategies (see, e.g., Litterman, 2003; Meucci, 2005) are also fixed strategies, any improvement over fixed strategies is of practical importance, which might be one of the reasons that technical analysis is widely used.[3]

However, because the MA, as a simple filter of the available information on the stock price, disregards any information on predictive variables, trading strategies related to the MA must be in general dominated by the optimal dynamic strategy, which optimally uses all available information on both the stock price and predictive variables. An argument in favor of the MA could be that the optimal dynamic strategy is difficult for investors at large to implement due to the difficulty of model identification as well as the cost of collecting and processing information. It is not easy to find reliable predictive variables, and their observations at desired time frequencies are not readily available in real time. This gives rise to the problem of predictability uncertainty in practice. In the presence of such uncertainty, Gennotte (1986), Barberis (2000), and Xia (2001), among others, show that the optimal dynamic strategy depends on optimal learning about the unknown parameters of the model and that, in turn, depends on the investor's prior on the parameters. In the context of the Xia (2001) model, we find that, with the use of the MA rule, one can outperform the optimal dynamic trading strategy when the priors are reasonable and yet not too informative. This seems due to the fact that the MA rule is less model dependent, and so it is more robust to the choice of underlying predictive variables.

Furthermore, the usefulness of the MA rule is more apparent when uncertainty exists about which model truly governs the stock price. In the real world, the true model is unknown to all investors. But for a wide class of plausible candidates of the true model, the optimal MA can be estimated easily, while the optimal trading strategy relies on a complete specification of the true model. When the wrong model is used to derive the optimal trading strategy, we show that the estimated optimal MA outperforms it substantially.

In typical applications, one usually chooses some ex ante value as the lag length of the MA. The question of using the optimal lag has been done only by trial and error, and only for the pure MA strategy that takes an

[1] Some academics take a strong view against technical analysis. For example, in his influential book, Malkiel (1981, p. 139) says that, "technical analysis is anathema to the academic world."

[2] As time passes, the average price is always computed based on its current price and on those in the most recent L periods. Hence, the average is called the MA.

[3] Behaviorial reasons, such as limited attention and optimal learning with limited resources, could explain the use of simple technical rules in practice, in addition to the rational reasons explored in this paper.

all-or-nothing allocation. Because this allocation itself is suboptimal, the associated optimal lag is suboptimal, too. The asset allocation perspective provided here not only solves the optimal stock allocation problem for both the pure MA and its optimal combination with the fixed rules, but also determines the optimal lag of the MA. We find that the fixed rules in conjunction with the MA are fairly insensitive to the use of the optimal lags, while the optimal generalized MA (GMA) is not.

The paper is organized as follows. In Section 2, we provide a literature review of the studies on technical analysis that are related to the current paper. In Section 3, we provide mainly our theoretical results. First, we outline the asset allocation model and investment strategies with the use of the MA. Second, we solve the optimal strategies explicitly in the log-utility case and obtain both the approximately analytical solutions in the power-utility case. Third, we analyze the strategies when there is either parameter uncertainty or model uncertainty. Finally, we explore the optimal choice of the MA lag length. In Section 4, we provide an empirical illustration on the performance of the strategies in calibrated models, and we conclude in Section 5.

2. Literature review

Technical analysis claims the ability to forecast the future direction of asset prices through the study of past market data. According to Nison (1991, p. 13), among the first and famous technicians who use past prices to predict future price movements is the legendary speculator Munehisa Homma, who amassed a huge fortune in the rice market in the 1700s in Japan and whose techniques evolved into what is known today as the candlestick patterns. In the United States, the Dow Theory, developed by Charles Dow and refined by William Peter Hamilton in the 1800s, asserts that the stock market moves in certain phases with predictable patterns. While the classic book by Murphy (1986) summarizes the Dow Theory and various other technical indicators, a growing and large literature has emerged on new techniques of technical analysis due to the wide availability of data and computing power (see, e.g., Covel, 2005; Kirkpatrick and Dahlquist, 2006). While technical analysts today could employ trading rules based, for example, on various price transformations and other market statistics, such as the relative strength index, cycles, and momentum oscillators, the MAs are the most popular and simple rules.

Cowles (1933), who seems to be the first to conduct an empirical study of technical analysis that is published in an academic journal, finds that Hamilton's forecasts based on the Dow Theory over the period of 1904 and 1929 are successful only 55% of the time. Subsequent studies on technical analysis are few until in the 1960s, when Fama and Blume (1966) show that common filter rules are not profitable based on daily prices of 30 individual securities in the Dow Jones Industrial Average (DJIA) over 1956–1962. Similar conclusion is also reached by Jensen and Benington (1970) in their study of relative strength systems. These empirical findings have perhaps prompted Fama (1970) to propose the well-known efficient market hypothesis that market prices reflect all available information so that no abnormal returns can be made with historical price and other market data.

The market efficiency was interpreted, in the earlier years by many, as a random walk model for the stock price. For any technical trading rule to be profitable, the stock return must be predictable, and so the use of the random walk model rules out any value of technical analysis. However, Lo and MacKinlay (1988) provide a variance ratio specification test that completely rejects the random walk model, supporting studies, such as Fama and Schwert (1977) and Campbell (1987), that various economic variables can forecast stock returns due to time-varying risk premiums. A huge literature is available on stock predictability, recent examples of which are Ferson and Harvey (1991), Lo and MacKinlay (1999), Goyal and Welch (2003), and Ang and Bekaert (2006). Current studies, such as Campbell and Thompson (2008), Cochrane (2008), Rapach, Strauss, and Zhou (2009) provide further evidence even on out-of-sample predictability. In addition, various asset pricing anomalies, for which Schwert (2003) provides an excellent survey, suggest predictable patterns of the stock returns. The predictability of stock returns allows for the possibility of profitable technical rules.

Brock, Lakonishok, and LeBaron (1992) provide strong evidence on the profitability of technical trading. With robust statistical tests, they find that simple trading rules, based on the popular MAs and range breakout, outperform the market over the 90 year period prior and up to 1987 based on daily data on DJIA. Moreover, in their comprehensive study of applying both kernel estimators and automated rules to hundreds of individual stocks, Lo, Mamaysky, and Wang (2000) find that technical analysis has added value to the investment process based on their novel approach comparing the distribution conditional on technical patterns, such as head-and-shoulders and double-bottoms, with the unconditional distribution. In contrast to the equity markets, the results in foreign exchange markets are generally much stronger. For example, LeBaron (1999) and Neely (2002), among others, show that substantial gains are made with the use of MAs and the gains are much larger than those in the stock market. Moreover, Gehrig and Menkhoff (2006) argue that technical analysis today is as important as fundamental analysis to currency mangers.

Statistically, though, it is difficult to show the true effectiveness of technical trading rules because of a data-snooping bias (see, e.g., Lo and MacKinlay, 1990), which occurs when a set of data is used more than once for the purpose of inference and model selection. In its simplest form, rules that are invented and tested by using the same data set are likely to exaggerate their effectiveness. Accounting for the data-snooping bias, for example, Sullivan, Timmermann, and White (1999) show via bootstrap that the Brock, Lakonishok, and LeBaron results are much weakened. Using generic algorithms, Allen and Karjalainen (1999) find little profitability in technical trading. One could then argue that a bootstrap is subject to specification bias and that generic algorithms can be inadequate due to inefficient ways of learning. In any case, it appears that the statistical debate on the effectiveness of technical analysis is unlikely to get settled soon.

Theoretically, few studies explain why technical analysis has value under certain conditions. In a two-period model with third period consumption, Brown and Jennings (1989) show that rational investors can gain from forming expectations based on historical prices. In an equilibrium model in which the volume also plays a role, Blume, Easley, and O'Hara (1994) show that traders who use information contained in market statistics do better than traders who do not. In a model of information asymmetry, Grundy and Kim (2002) also find value of using technical analysis.[4] However, to our knowledge, no theoretical studies are closely tied to the conventional use of technical analysis, and no studies calibrate the model to data to provide insights on the realistic use of technical analysis in practice. The exploratory study here attempts to fill this gap of the literature. In so doing, we study the classic asset allocation problem and examine how technical analysis, especially the MA, can be optimally used to add value to the investment process.

3. The model and analytic results

In this section, we present the model and analytical results for various cases. To focus on ideas and intuition, we motivate the framework and only explain the main results, while leaving the derivations of all major formulas and propositions to Appendix A.

3.1. The model and investment strategies

For simplicity, we consider a two-asset economy in which a riskless bond pays a constant rate of interest r, and a risky stock represents the aggregate equity market. Because of the ample evidence on the predictability of stock returns[5] we follow Kim and Omberg (1996) and Huang and Liu (2007), among others, and assume the following dynamics for the cum-dividend stock price S_t:

$$\frac{dS_t}{S_t} = (\mu_0 + \mu_1 X_t)\,dt + \sigma_s\,dB_t \tag{1}$$

and

$$dX_t = (\theta_0 + \theta_1 X_t)\,dt + \sigma_x\,dZ_t, \tag{2}$$

where $\mu_0, \mu_1, \sigma_s, \theta_0, \theta_1$, and σ_x are parameters; X_t is a predictive variable; and B_t and Z_t are standard Brownian motions with correlation coefficient ρ; θ_1 has to be negative to make X_t a mean-reverting process. The model is a special case of the general model of Merton (1992). In discrete-time, it is the well-known predictive regression model (e.g., Stambaugh, 1999).

Given an initial wealth W_0 and an investment horizon T, the standard allocation problem of an investor is to choose a portfolio strategy ξ_t to maximize his expected utility of wealth,

$$\max_{\xi_t} \mathrm{E}[u(W_T)], \tag{3}$$

subject to the budget constraint

$$\frac{dW_t}{W_t} = r\,dt + \xi_t(\mu_0 + \mu_1 X_t - r)\,dt + \xi_t \sigma_s\,dB_t. \tag{4}$$

The solution to this problem is the optimal trading strategy. In general, this strategy is a function of time and the associated state variables. We refer to it as the optimal dynamic strategy, because it varies with time and states.

In this paper, we assume the power-utility

$$u(W_T) = \frac{W_T^{1-\gamma}}{1-\gamma}, \tag{5}$$

where γ is the investor's risk-aversion parameter. In this case, the optimal dynamic strategy is known (see, e.g., Kim and Omberg, 1996; Huang and Liu, 2007) and is given by

$$\xi_t^* = \frac{\mu_0 + \mu_1 X_t - r}{\gamma\sigma_s^2} + \frac{(1-\gamma)\rho\sigma_x}{\gamma\sigma_s}[\chi(t) + \zeta(t)X_t], \tag{6}$$

where $\chi(t)$ and $\zeta(t)$ satisfy the following ordinary differential equations:

$$\dot{\chi}(t) + a_1\zeta(t)\chi(t) + \tfrac{1}{2}a_2\chi(t) + a_4\zeta(t) + a_5 = 0 \tag{7}$$

and

$$\dot{\zeta}(t) + a_1\zeta^2(t) + a_2\zeta(t) + a_3 = 0, \tag{8}$$

with

$$a_1 = \frac{(1-\gamma)^2}{\gamma}\rho^2\sigma_x^2 + (1-\gamma)\sigma_x^2, \tag{9}$$

$$a_2 = 2\left(\frac{1-\gamma}{\gamma}\frac{\mu_1}{\sigma_s}\rho^2\sigma_x^2 + \theta_1\right), \tag{10}$$

$$a_3 = \frac{1}{\gamma}\left(\frac{\mu_1}{\sigma_s}\right)^2, \tag{11}$$

$$a_4 = \frac{1-\gamma}{\gamma}\frac{\mu_0 - r}{\sigma_s}\rho\sigma_x + \theta_0, \tag{12}$$

$$a_5 = \frac{\mu_1(\mu_0 - r)}{\gamma\sigma_s^2}, \tag{13}$$

and the terminal conditions $\chi(T) = \zeta(T) = 0$.

The assumption that stock returns are independently and identically distributed (iid) over time has played a

[4] In addition, behavioral models, such as those reviewed by Shleifer (2000) and Shefrin (2008), offer support to technical analysis by theorizing certain predictable patterns of the market.

[5] Kandel and Stambaugh (1996), Barberis (2000), and Huang and Liu (2007) are examples of studies on portfolio choice under predictability.

major role in finance. It was the basis for much of the earlier market efficiency arguments, although it was known later as only a sufficient condition. Nevertheless, some of the most popular investment strategies and theoretical models are based on this assumption. Under the iid assumption, the optimal strategy is

$$\xi^*_{\text{fix1}} = \frac{\mu_s - r}{\gamma\sigma_s^2}, \tag{14}$$

where μ_s is the long-term mean of the stock return. This strategy invests a fixed or constant proportion of wealth, ξ^*_{fix1}, into the stock all the time. In discrete-time, this is the familiar suggestion of the Markowitz (1952) mean–variance framework and the Tobin (1958) two-fund separation theorem.[6] The strategy is one of the most important benchmark models used in practice today (see, e.g., Litterman, 2003; Meucci, 2005). Because of it, passive index investments have become increasingly popular (Rubinstein, 2002). Theoretically, the allocation rule ignores any time-varying investment opportunities and is clearly not optimal once the iid assumption is violated. A likely practical motivation for its wide use is as follows. Although stock returns are predictable, the predictability is small and uncertain. It could be costly for a small investor to collect news and reports about X_t whose costs could outweigh the benefits. As a result, the investor could simply follow a fixed rule even though there is a small degree of predictability.

The fixed rule ξ^*_{fix1} ignores any predictability completely. An interesting question is, then, whether one can obtain yet another fixed rule that accounts for the predictability. In other words, how should the investor invest his money when he knows the true predictive process but not the state variables? Mathematically, this amounts to solving the optimal allocation problem by restricting ξ_t to a constant. The solution is analytically obtained as

$$\xi^*_{\text{fix2}} = \frac{\mu_s - r}{\gamma\sigma_s^2 - (1-\gamma)(\mu_1^2 A + 2\mu_1\sigma_s B)}, \tag{15}$$

where

$$A = \frac{\sigma_x^2}{\theta_1^2}\left(1 + \frac{1 - e^{\theta_1 T}}{\theta_1 T}\right) \tag{16}$$

and

$$B = \frac{\rho\sigma_x}{\theta_1}\left(\frac{e^{\theta_1 T} - 1}{\theta_1 T} - 1\right). \tag{17}$$

For $\gamma = 1$, this optimal constant strategy is equal to ξ^*_{fix1}. In other words, for investors with log-utility, the optimal fixed strategy remains the same as before, even though the stock returns are predictable, a fact we can explain largely by the myopic behavior dictated by the log-utility function. For $\gamma > 1$, however, there is an adjustment in the denominator of Eq. (15). In general, the adjustment can be either positive or negative.

Let $L > 0$ be the lag or lookback period. A continuous-time version of the MA of the stock price at any time t is defined as

$$A_t = \frac{1}{L}\int_{t-L}^{t} S_u\, du, \tag{18}$$

i.e., the average price over time period $[t-L, t]$. The simplest MA trading rule is the following stock allocation strategy:

$$\eta_t = \eta(S_t, A_t) = \begin{cases} 1 & \text{if } S_t > A_t, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

This is well defined when $t > L$ and can be taken as zero or as another fixed constant when $t \leq L$.[7] This standard (pure) MA rule is a market timing strategy that shifts investments between cash and stock. Almost all existing studies on the MA strategy take a 100% position in stock or nothing, i.e., the portfolio weight (on the stock) is η_t. This is clearly not optimal for two reasons. First, the MA rule should in general be a function of the risk-aversion parameter γ. Intuitively, γ reflects the investor's tolerance to stock risk, and it has to enter the allocation decision as is the case for the earlier optimal fixed strategies. Second, the degree of predictability must matter. The more predictable the stock, the more reliable the MA rule and hence the more allocation to the stock.

Other than the pure MA rule, we also consider the following GMA rule,

$$\text{GMA}(S_t, A_t, \gamma) = \xi_{\text{fix}} + \xi_{\text{mv}} \cdot \eta(S_t, A_t), \tag{20}$$

where ξ_{fix} and ξ_{mv} are constants. This trading strategy is a linear combination of a fixed strategy and a pure MA strategy. It consists of all the previous strategies as special cases. For example, ξ^*_{fix1} is obtained by setting $\xi_{\text{fix}} = \xi^*_{\text{fix1}}$ and $\xi_{\text{mv}} = 0$, and η_t is obtained by setting $\xi_{\text{fix}} = 0$ and $\xi_{\text{mv}} = 1$.

Three interesting questions are associated with the GMA rule. First, what is the optimal choice of ξ_{fix} and ξ_{mv}, and how well does it perform compared with other fixed strategies? Second, with ξ_{fix} being equal to either ξ^*_{fix1} or ξ^*_{fix2}, is the optimal choice of ξ_{mv} zero or not? This indicates whether there is a gain in the expected utility when the fixed strategy is used in conjunction with the MA rule. Third, imposing $\xi_{\text{fix}} = 0$, what is the optimal choice of ξ_{mv}? This indicates the optimal amount of investment based purely on the MA trading signal. If $\xi_{\text{mv}} = 1$, the usual application of the MA with 100% stock allocation is optimal. However, as easily seen from our analysis below, the optimal value of ξ_{mv} is unlikely to be equal to one. These three questions are answered first analytically for the log-utility and then approximately analytically for the power-utility.

Analytically, the distribution of the arithmetic MA A_t is very complex and difficult to analyze. However, the

[6] See Ingersoll (1987) or Back (2006) for an excellent textbook exposition.

[7] In practice, the MA rule is computed based on ex-dividend prices, whose impact is analyzed in Section 4. Its starting time also needs be specified, which is discussed in Appendix A.

geometric MA,

$$G_t = \exp\left(\frac{1}{L}\int_{t-L}^{t} \log(S_u)\,du\right), \tag{21}$$

is tractable to allow explicit solutions. In addition, as shown in our simulations, little performance differences emerge in our main results with the use of either averages. Henceforth, we focus our analysis on $\mathrm{GMA}(S_t, G_t, \gamma)$, i.e., the GMA strategy based on the geometric average.

3.2. Explicit solutions under log-utility

In this subsection, we provide the explicit solutions to the optimal GMA strategies and compare them analytically with both the optimal fixed and the optimal dynamic allocations.

The wealth process corresponding to the GMA is

$$\frac{dW_t}{W_t} = [r + \mathrm{GMA}\cdot(\mu_0 + \mu_1 X_t - r)]\,dt + \mathrm{GMA}\cdot\sigma_s\,dB_t \tag{22}$$

and, hence, assuming $T>L$, we have

$$\begin{aligned}\log W_T = {} & \log W_0 + rT \\ & + \int_0^L dt\left[\xi^*_{\mathrm{fix}1}\left(\mu_0 + \mu_1 X_t - r - \frac{\sigma_s^2}{2}\xi^*_{\mathrm{fix}1}\right)\right] \\ & + \int_L^T dt\left[\xi_{\mathrm{fix}}\left(\mu_0 + \mu_1 X_t - r - \frac{\sigma_s^2}{2}\xi_{\mathrm{fix}}\right)\right] \\ & + \xi_{\mathrm{mv}}\mu_1\int_L^T dt\hat{X}_t\eta_t + \int_L^T dt\left[\xi_{\mathrm{mv}}(\mu_0 + \mu_1\bar{X} - r)\right. \\ & \left. - \frac{\sigma_s^2}{2}\xi_{\mathrm{mv}}^2 - \sigma_s^2\xi_{\mathrm{fix}}\xi_{\mathrm{mv}}\right]\eta_t \\ & + \sigma_s\int_L^T(\xi_{\mathrm{fix}} + \xi_{\mathrm{mv}}\eta_t)\,dB_t,\end{aligned} \tag{23}$$

where $\hat{X}_t = X_t - \bar{X}$ with $\bar{X} = -\theta_0/\theta_1$. Assume that X_t is stationary, and it starts from its steady state distribution.[8] Then, the expected log-utility is

$$\begin{aligned}U_{\mathrm{GMA}} = \mathrm{E}\log W_T = {} & \log W_0 + rT + \frac{(\mu_0 + \mu_1\bar{X} - r)^2}{2\sigma_s^2}L \\ & + \int_L^T dt\xi_{\mathrm{fix}}\left[\mu_0 + \mu_1\bar{X} - r - \frac{\sigma_s^2}{2}\xi_{\mathrm{fix}}\right] \\ & + \int_L^T dt\xi_{\mathrm{mv}}\mu_1\mathrm{E}[\hat{X}_t\eta_t] \\ & + \int_L^T dt[\xi_{\mathrm{mv}}(\mu_0 + \mu_1\bar{X} - r) \\ & \left. - \frac{\sigma_s^2}{2}\xi_{\mathrm{mv}}^2 - \sigma_s^2\xi_{\mathrm{fix}}\xi_{\mathrm{mv}}\right]\mathrm{E}[\eta_t].\end{aligned} \tag{24}$$

To solve the optimization problem, let

$$b_1 \equiv \mathrm{E}[\hat{X}_t\eta(S_t, G_t)] \tag{25}$$

and

$$b_2 \equiv \mathrm{E}[\eta(S_t, G_t)], \tag{26}$$

where b_1 is the covariance between X_t and the MA strategy η_t and b_2 is the probability of $S_t > G_t$ at any given time. It can be shown that

$$b_1 = \mathrm{E}\hat{X}_t\eta(S_t, G_t) = \frac{C_{12}^Z}{\sqrt{C_{22}^Z}}\mathrm{N}'\left(-\frac{m_2^Z}{\sqrt{C_{22}^Z}}\right) \tag{27}$$

$$b_2 = \mathrm{E}\eta(S_t, G_t) = \mathrm{N}\left(\frac{m_2^Z}{\sqrt{C_{22}^Z}}\right), \tag{28}$$

where

$$C_{12}^Z = \left(\frac{\mu_1\sigma_x^2}{2\theta_1^2} - \frac{\sigma_x\sigma_s\rho}{\theta_1}\right)\left(1 - \frac{e^{\theta_1 L} - 1}{\theta_1 L}\right), \tag{29}$$

$$\begin{aligned}C_{22}^Z = {} & \left(\sigma_s^2 + \frac{\mu_1^2\sigma_x^2}{\theta_1^2} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}\right)\frac{L}{3} + \left(\frac{\mu_1^2\sigma_x^2}{2\theta_1^3} - \frac{\mu_1\sigma_x\sigma_s\rho}{\theta_1^2}\right) \\ & \times\left[1 - \frac{2}{(\theta_1 L)^2}(1 - e^{\theta_1 L} + \theta_1 L e^{\theta_1 L})\right],\end{aligned} \tag{30}$$

$$m_2^Z = \left(\mu_0 + \mu_1\bar{X} - \frac{\sigma_s^2}{2}\right)\frac{L}{2}, \tag{31}$$

and $\mathrm{N}(\cdot)$ and $\mathrm{N}'(\cdot)$ are the distribution and density functions of the standard normal random variable, respectively. Because X_t starts from its steady state distribution, b_1 and b_2 are independent of time t. Therefore, the expected log-utility of Eq. (24) becomes

$$\begin{aligned}U_{\mathrm{GMA}} = \mathrm{E}\log W_T = {} & \log W_0 + rT + \frac{(\mu_0 + \mu_1\bar{X} - r)^2}{2\sigma_s^2}L \\ & + \xi_{\mathrm{fix}}\left[\mu_0 + \mu_1\bar{X} - r - \frac{\sigma_s^2}{2}\xi_{\mathrm{fix}}\right](T - L) \\ & + \xi_{\mathrm{mv}}\mu_1 b_1(T - L) + \left[\xi_{\mathrm{mv}}(\mu_0 + \mu_1\bar{X} - r)\right. \\ & \left. - \frac{\sigma_s^2}{2}\xi_{\mathrm{mv}}^2 - \sigma_s^2\xi_{\mathrm{fix}}\xi_{\mathrm{mv}}\right]b_2(T - L).\end{aligned} \tag{32}$$

With these preparations, we are ready to answer the three questions raised earlier. In doing so, we assume that the investment horizon T is greater than or equal to the lag length L throughout. This assumption is clearly harmless.

3.2.1. Optimal GMA

On the question of finding the GMA strategy that combines a fixed rule with the MA, the results are given by Proposition 1.

Proposition 1. *In the class of strategies $\mathrm{GMA}(S_t, G_t, \gamma)$, the optimal choice of ξ_{fix} and ξ_{mv} under the log-utility is*

$$\xi^*_{\mathrm{fix}} = \frac{\mu_s - r}{\sigma_s^2} - \frac{\mu_1 b_1}{(1 - b_2)\sigma_s^2}, \tag{33}$$

$$\xi^*_{\mathrm{mv}} = \frac{\mu_1 b_1}{b_2(1 - b_2)\sigma_s^2}, \tag{34}$$

[8] Our goal here is to find the unconditionally optimal GMA rule. In other words, we solve in what follows the optimal allocation problem using the steady state distribution for X_0. See, e.g., Karatzas and Shreve, 1991, p. 358, for a general discussion on the steady state.

and the associated value function is

$$U^*_{\mathrm{GMA1}} = U^*_{\mathrm{fix1}} + \frac{\mu_1^2 b_1^2}{2b_2(1-b_2)\sigma_s^2}(T-L) \geqslant U^*_{\mathrm{fix1}}, \tag{35}$$

where U^*_{fix1} is the value function associated with ξ^*_{fix1}.

Proposition 1 says that the improvement over ξ^*_{fix1} is always positive by combining a suitable fixed strategy with the MA one unless $\mu_1 = 0$. In the case of $\mu_1 = 0$, the stock return is unpredictable, and the fixed strategy ξ^*_{fix1} is optimal already. The point is that ξ^*_{fix1} is not optimal in general, and so the MA rule can help to gain in expected utility with the combination of another fixed strategy. In the log-utility case, $\xi^*_{\mathrm{fix2}} = \xi^*_{\mathrm{fix1}}$. Hence, Proposition 1 applies to ξ^*_{fix2} as well, and ξ^*_{fix1} is the only fixed strategy to compare with.

It is interesting to observe that

$$\xi^*_{\mathrm{fix}} + (b_2 \xi^*_{\mathrm{mv}}) = \xi^*_{\mathrm{fix1}}. \tag{36}$$

If the predictive variable X_t is positively related to the stock market with $\mu_1 > 0$ and $\rho > 0$, the investor invests less than the standard fixed strategy by the amount of $b_2 \xi^*_{\mathrm{mv}}$ because $0 < b_2 < 1$ and $\xi^*_{\mathrm{mv}} > 0$. Once the trend is up, as suggested by the MA rule, the investor is more aggressive than the fixed strategy by investing an extra amount of $(1-b_2)\xi^*_{\mathrm{mv}}$. This is consistent with the intuition that one should take advantage of the predictability of the stock market once it is detected by the MA rule.

If one strategy outperforms another over horizon T, it must continue to do so over a longer time. Hence, $U^*_{\mathrm{GMA1}} - U^*_{\mathrm{fix1}}$ must be an increasing function of T. What is striking here is that this relation is in fact linear in T in the log-utility case, because b_1, b_2, μ_1, and σ_s are all horizon-independent parameters.

Proposition 1 also makes possible an analytical comparison between GMA1 and the optimal dynamic strategy. Under the log-utility, the optimal dynamic rule of Eq. (6) is the same as the myopic rule

$$\xi^*_{\mathrm{opt}} = \frac{\mu_0 + \mu_1 X_t - r}{\sigma_s^2}. \tag{37}$$

By substituting this optimal rule into the wealth process, we obtain the optimal utility

$$U^*_{\mathrm{opt}} = U^*_{\mathrm{fix}} + \frac{1}{2}\frac{\mu_1^2 \mathrm{E}\hat{X}_t^2}{\sigma_s^2} T. \tag{38}$$

Based on the value functions in both cases, we have

$$U^*_{\mathrm{opt}} - U^*_{\mathrm{GMA1}} \geqslant \frac{\mu_1^2}{2\sigma_s^2}\left[\mathrm{E}\hat{X}_t^2 - \frac{b_1^2}{b_2(1-b_2)}\right](T-L). \tag{39}$$

Recalling that $b_1 = \mathrm{E}\hat{X}_t\eta$ and $b_2 = \mathrm{E}\eta$, we have $\mathrm{var}(\eta) = \mathrm{E}\eta^2 - (\mathrm{E}\eta)^2 = b_2(1-b_2)$, and hence

$$\frac{b_1^2}{b_2(1-b_2)} = \frac{(\mathrm{E}\hat{X}_t\eta)^2}{\mathrm{var}(\eta)} = \frac{(\mathrm{cov}(\hat{X}_t,\eta))^2}{\mathrm{var}(\eta)} \leqslant \frac{\mathrm{E}(\hat{X}_t^2)\mathrm{var}(\eta)}{\mathrm{var}(\eta)} = \mathrm{E}\hat{X}_t^2. \tag{40}$$

Therefore, the right-hand side of Eq. (39) is always positive, as it must be, because U^*_{opt} is the expected utility under the optimal dynamic strategy. It is seen that the smaller the σ_x^2, the smaller the difference. In other words, the less volatile the predictive variable, the closer the GMA1 to the optimal strategy. However, it should also be noted that, as σ_x^2 gets smaller, b_1 also gets closer to zero, i.e., the MA component becomes smaller, too.

3.2.2. Combining a fixed rule with MA

Now we consider whether the MA strategy can be used in conjunction with ξ^*_{fix1} to add value. To address this issue, we need to solve the earlier optimization by imposing the constraint that $\xi_{\mathrm{fix}} = \xi^*_{\mathrm{fix1}}$. In this case, we present Proposition 2.

Proposition 2. *In the class of strategies $\mathrm{GMA}(S_t, G_t, \gamma)$ with ξ_{fix} being set at ξ^*_{fix1}, the optimal choice of ξ_{mv} under the log-utility is*

$$\xi^*_{\mathrm{mv}} = \frac{\mu_1 b_1}{b_2 \sigma_s^2}, \tag{41}$$

and the associated value function is

$$U^*_{\mathrm{GMA2}} = U^*_{\mathrm{fix1}} + \frac{\mu_1^2 b_1^2}{2b_2\sigma_s^2}(T-L) \geqslant U^*_{\mathrm{fix1}}, \tag{42}$$

where U^*_{fix1} is the value function associated with ξ^*_{fix1}.

As for U^*_{GMA1}, U^*_{GMA2} is at least as large as U^*_{fix1}. When there is predictability, U^*_{GMA2} is clearly strictly larger than U^*_{fix1}, implying that the MA rule helps to improve the expected utility and does so strictly as long as the stock return is predictable.

An interesting observation is that ξ^*_{mv} in Proposition 2 differs from that in Proposition 1 by only a factor of $1-b_2$ in the denominator. Because $0 < b_2 < 1$, ξ^*_{mv} is smaller now in absolute value. This is expected. Because ξ^*_{fix} is set at ξ^*_{fix1}, the risk exposure to the stock market is relatively higher already as $\xi^*_{\mathrm{fix1}} > \xi^*_{\mathrm{fix}}$. Hence, when the MA rule detects an upward trend in the market, the investor acts more aggressively than ξ^*_{fix1}, but less aggressively than before. Finally, it is seen that

$$U^*_{\mathrm{GMA2}} = U^*_{\mathrm{GMA1}} - \frac{\mu_1^2 b_1^2}{2(1-b_2)\sigma_s^2}(T-L) \leqslant U^*_{\mathrm{GMA1}} \leqslant U^*_{\mathrm{opt}}. \tag{43}$$

While the second inequality is obvious, the first inequality should be true, too. The fixed component of GMA1 is optimally chosen, and hence its performance must be better than the GMA strategy with that component set at ξ^*_{fix1}.

3.2.3. Optimal pure MA

A standard or pure MA rule is a market timing strategy that shifts money between cash and risky assets. Existing studies provide no guidance as to how much one should optimally invest in the stock even if one believes it is in an up-trend as signaled by the MA rule. Clearly, a 100% investment in the stock market is not optimal from a utility maximization point of view. Here we solve the optimal amount explicitly.

Proposition 3. *In the class of strategies $\mathrm{GMA}(S_t, G_t, \gamma)$ with restriction $\xi_{\mathrm{fix}} = 0$, the optimal choice of ξ_{mv} under* *the log-utility is*

$$\xi^*_{\text{mv}} = \frac{\mu_s - r}{\sigma_s^2} + \frac{\mu_1 b_1}{b_2\sigma_s^2}, \tag{44}$$

and the associated value function is

$$U^*_{\text{GMA3}} = U^*_{\text{fix1}} + \frac{(\mu_1 b_1 + (\mu_s - r)b_2)^2 - (\mu_s - r)^2 b_2}{2b_2\sigma_s^2}(T - L), \tag{45}$$

which can be either greater or smaller than U^*_{fix1}, the value function associated with ξ^*_{fix1}.

Consistent with our intuitive reasoning in the Introduction, Proposition 3 says that, if an all-or-nothing investment strategy is taken based on the MA, the optimal stock allocation is unlikely to be 100%. Recognizing that 100% is not optimal, one could suggest a two-step approach for making use of the MA signal. In the first step, one determines the stock allocation, say ξ^*_{fix1}, based on a standard fixed allocation model and then, in the second step, apply this in the market-timing decision: invest that amount into the stock if MA signals a buy, and nothing otherwise. Eq. (44) says that this fixed amount differs from ξ^*_{mv} in general, and hence the decision is suboptimal, too. The intuition is that one should invest more than that fixed amount if an up trend is detected, and less if there is a down trend.

Proposition 3 also says that whether or not the pure MA strategy can outperform the fixed strategy depends on particular parameter values. It can be verified that if

$$\mu_s - r < \frac{\mu_1 b_1}{\sqrt{b_2} - b_2}, \tag{46}$$

the relation about the risk premium, is satisfied, then the pure MA strategy does yield a higher expected utility than the fixed strategy ξ^*_{fix1}.[9] However, with reasonable parameters calibrated from data, this condition is not satisfied. It implies that the optimal pure MA strategy usually performs worse than the simple fixed strategy. Our later simulations show that, the pure MA strategy and its common analogues always perform the worst. Hence, if the MA rule is to be of any value to investors, it must be used wisely and in conjunction with the fixed strategies demonstrated by Propositions 1 and 2.

3.3. Analytic solutions under power-utility

In this subsection, we extend our earlier analysis to the power-utility case. First, we provide first-order accurate analytical solutions to the fixed strategies combined optimally with the MA. The analytical solutions provide insight on the role played by an investor's risk aversion. Second, we derive second-order accurate analytical solutions to the strategies that are important for computing their performance under the power-utility.

3.3.1. First-order approximate solutions

In the power-utility case, the complexity of the utility function precludes us from deriving exact analytical solutions to those trading strategies examined earlier. Nevertheless, we can obtain first-order analytical approximations. The solutions reveal how the trading strategies are affected by γ, the investor's risk aversion.

By approximating $\int_0^T X_t\,dt$, $\int_0^T X_t\eta_t\,dt$ and $\int_0^T \eta_t\,dt$ with their mean values, we can write the expected utility under the GMA as

$$\begin{aligned} U_{\text{GMA}}(\gamma) \approx{} & \frac{(W_0\exp(rT))^{1-\gamma}}{1-\gamma} \\ & \cdot \exp\left\{(1-\gamma)T\left[\xi_{\text{fix}}(\mu_0 + \mu_1\bar{X} - r) - \frac{\gamma\sigma_s^2}{2}\xi_{\text{fix}}^2\right.\right. \\ & + \xi_{\text{mv}}\mu_1 \mathrm{E}\left[\hat{X}_t\eta_t\right] + \left[\xi_{\text{mv}}(\mu_0 + \mu_1\bar{X} - r)\right. \\ & \left.\left.\left. - \frac{\gamma\sigma_s^2}{2}\xi_{\text{mv}}^2 - \gamma\sigma_s^2\xi_{\text{fix}}\xi_{\text{mv}}\right]\mathrm{E}\eta_t\right]\right\}. \end{aligned} \tag{47}$$

Optimizing this approximated utility function, we obtain

$$\text{GMA}(S_t, G_t, \gamma) = \frac{1}{\gamma}\text{GMA}(S_t, G_t, 1). \tag{48}$$

This says that the optimal GMA rules in the $\gamma \neq 1$ case is simply a scale of those in the log-utility case. Hence, much of the qualitative results obtained in the log-utility case carry over to the power-utility case, with accuracy up to the first-order approximation.

For example, the GMA1 strategy in the power-utility case is still of the earlier form, but with

$$\xi^*_{\text{fix}} = \frac{\mu_s - r}{\gamma\sigma_s^2} - \frac{\mu_1 b_1}{\gamma(1-b_2)\sigma_s^2} \tag{49}$$

and

$$\xi^*_{\text{mv}} = \frac{\mu_1 b_1}{\gamma b_2(1-b_2)\sigma_s^2}. \tag{50}$$

This says that we simply scale down the stock investment by $1/\gamma$ when the investor is more risk-averse than the log-utility case. The same conclusion also holds for other strategies. This scaling corresponds precisely to the way by which the usual fixed strategy is adjusted when the investor's preference changes from the log- to the power-utility. In particular, the optimal pure MA rule depends on γ. However, one should keep in mind that the simple inverse dependence on γ here is not exact, but only approximate with first-order accuracy.

3.3.2. Second-order approximate solutions

While the previous approximate solutions make apparent the role of γ, they are not accurate enough in simulations for measuring the true performance of the optimal GMA strategies, which are analytically unavailable. One could propose a numerical method, such as simulation, to compute the optimal GMA strategies, but this is feasible only for a given S_t, G_t, and t. To evaluate the performance of these strategies, however, we need to

[9] To appreciate the intuition behind the condition, we note that the denominator of the right-hand side of the inequality is dominated by 0.25. Therefore, a sufficient condition for pure MA strategy to outperform a fixed rule is $\mu_1 b_1 > 4(\mu_s - r)$, which means that, when predictability is stronger, the MA strategy is more likely to dominate the fixed rule. Similarly, if the equity premium is not too large, the MA strategy is more likely to dominate.

compute the optimal GMA strategies at hundreds and thousands of draws of S_t and G_t and time t. Therefore, due to the curse of dimensionality, it is not possible to evaluate the performance of the optimal GMA strategies numerically without efficiently determining the strategies in the first place. To resolve this problem, we now derive alternative analytical solutions to the strategies. These are more complex than the earlier ones but are accurate to the second-order. As a compromise, they are taken as the true strategies. Simulations are used to evaluate their performances.

Instead of ignoring the second-order terms of the random variables in Eq. (23), we approximate them by a Gaussian process that match both the first and second moments. Then, the power-utility,

$$U(\gamma) = \frac{1}{1-\gamma}\mathrm{E}[W_T^{1-\gamma}] = \frac{1}{1-\gamma}\mathrm{E}[\exp((1-\gamma)\log W_T)], \quad (51)$$

can be approximated by

$$\begin{aligned} U(\gamma) = {} & \frac{(W_0\exp(rT))^{1-\gamma}}{1-\gamma}U_{\text{fix}}(\xi_{\text{fix}}) \\ & \exp\Big\{(1-\gamma)\xi_{\text{mv}}\mathrm{E}[C_T + D_T + y(\xi_{\text{fix}},\xi_{\text{mv}})F_T] \\ & +\frac{1}{2}(1-\gamma)^2\xi_{\text{mv}}^2\mathrm{var}[C_T + D_T + y(\xi_{\text{fix}},\xi_{\text{mv}})F_T] \\ & +(1-\gamma)^2\xi_{\text{fix}}\xi_{\text{mv}}\mathrm{cov}(A_T + B_T, C_T + D_T + yF_T)\Big\}, \quad (52) \end{aligned}$$

where $U_{\text{fix}}(\xi_{\text{fix}})$ is the value function associated with a given fixed strategy ξ_{fix},

$$y(\xi_{\text{fix}},\xi_{\text{mv}}) = (\mu_0 + \mu_1\bar{X} - r) - \tfrac{1}{2}\sigma_s^2\xi_{\text{mv}} - \sigma_s^2\xi_{\text{fix}} \quad (53)$$

and

$$C_T = \mu_1\int_0^T \eta_t X_t\,dt, \quad (54)$$

$$D_T = \sigma_s\int_0^T \eta_t\,dB_t, \quad (55)$$

$$F_T = \int_0^T \eta_t\,dt, \quad (56)$$

$$A_T = \mu_1\int_0^T X_t\,dt, \quad (57)$$

$$B_T = \sigma_s\int_0^T dB_t. \quad (58)$$

Upon some further algebraic manipulation, we obtain the power-utility value function as

$$\begin{aligned} U(\gamma) = {} & \frac{(W_0\exp(rT))^{1-\gamma}}{1-\gamma}U_{\text{fix}}(\xi_{\text{fix}})\exp\{(1-\gamma)\xi_{\text{mv}}[\phi_0 \\ & + \phi_1\xi_{\text{mv}} + \phi_2\xi_{\text{mv}}^2 + \phi_3\xi_{\text{mv}}^3]\}, \quad (59) \end{aligned}$$

where

$$\begin{aligned} \phi_0 = {} & \mathrm{E}C_T + (\mu_0 + \mu_1\bar{X} - r - \sigma_s^2\xi_{\text{fix}})\mathrm{E}F_T \\ & + (1-\gamma)\xi_{\text{fix}}\mathrm{cov}(A_T + B_T, C_T + D_T \\ & + (\mu_0 + \mu_1\bar{X} - r - \sigma_s^2\xi_{\text{fix}})F_T), \quad (60) \end{aligned}$$

$$\begin{aligned} \phi_1 = {} & -\tfrac{1}{2}\sigma_s^2\mathrm{E}F_T + \tfrac{1}{2}(1-\gamma)\mathrm{var}(C_T + D_T \\ & + (\mu_0 + \mu_1\bar{X} - r - \sigma_s^2\xi_{\text{fix}})F_T) \\ & + (1-\gamma)\xi_{\text{fix}}\mathrm{cov}(A_T + B_T, -\tfrac{1}{2}F_T), \quad (61) \end{aligned}$$

$$\begin{aligned} \phi_2 = {} & (1-\gamma)\mathrm{cov}(C_T + D_T \\ & + (\mu_0 + \mu_1\bar{X} - r - \sigma_s^2\xi_{\text{fix}})F_T, \\ & -\tfrac{1}{2}\sigma_s^2F_T), \quad (62) \end{aligned}$$

and

$$\phi_3 = \frac{1}{2}(1-\gamma)\frac{\sigma_s^4}{4}\mathrm{var}(F_T). \quad (63)$$

Hence, for any given ξ_{fix}, we can solve the associated ξ_{mv}^*, which maximizes $U(\gamma)$, Eq. (59), as

$$\begin{aligned} \xi_{\text{mv}}^* = {} & -\frac{\phi_2}{4\phi_3} - \left[\frac{q + \sqrt{q^2 + 4p^3/27}}{2}\right]^{1/3} \\ & + \frac{p}{3}\left[\frac{q + \sqrt{q^2 + 4p^3/27}}{2}\right]^{-1/3}, \quad (64) \end{aligned}$$

where

$$p = \frac{\phi_1}{3\phi_3} - \frac{1}{3}\left(\frac{2\phi_2}{3\phi_3}\right)^{1/3} \quad (65)$$

and

$$q = \frac{\phi_0}{3\phi_3} - \frac{2}{27}\frac{\phi_0\phi_1\phi_2}{\phi_3^3} + \frac{2}{27}\left(\frac{2\phi_2}{3\phi_3}\right)^3. \quad (66)$$

If $\xi_{\text{fix}} = \xi_{\text{fix1}}^*$ or ξ_{fix2}^* or 0, we obtain the corresponding ξ_{mv}^* from Eq. (64) that yields the approximate optimal GMA strategies. For easier reference, we denote them as Fix1+MA, Fix2+MA, and PureMA, respectively. These three together with ξ_{fix1}^* and ξ_{fix2}^*, denoted as Fix1 and Fix2, consist of five strategies whose performances are examined in detail in Section 4.

Finally, we note two interesting cases in which our analysis can be extended to allow intermediate consumption. The first is to assume a complete market under the current power-utility. Based on Wachter (2002) and Liu (2007), the indirect utility with intermediate consumption is a weighted average of the indirect utility with terminal wealth only, and hence the portfolio policy is similar. However, because the complete market assumes a perfect correlation between the stock return and the predictive variable, which is unrealistic in our context, we omit the analysis here. The second case is to use the Epstein, Zin, and Weil or recursive utility, i.e., the stochastic differential utility in continuous-time. When the coefficient of the elasticity of intertemporal substitution is one, the consumption is a constant ratio of wealth, and hence the portfolio policy is the same with or without consumption. When the risk-aversion coefficient is one, the portfolio policy consists of the myopic one only, and consumption does not affect portfolio choice. Under the later condition, as shown by Campbell and Viceira (1999), the consumption affects the portfolio policy only through the hedging demand, which is proportional to the covariance between the predictive variable and the consumption–wealth ratio. Under both conditions, the optimal portfolio with the GMA remains the same, although the utility losses could be bigger due to early consumption.

3.4. Solutions under parameter uncertainty

In previous subsections, we follow the common assumption that an economic agent making an optimal financial decision knows the true parameters of the model. However, the decision maker rarely, if ever, knows the true parameters. In reality, model parameters have to be estimated, and different parameter estimates could provide entirely different results. This gives rise to the estimation risk associated with any trading strategy. In this subsection, we analyze the performance of various investment strategies under such parameter uncertainty.

One remarkable feature of the pure MA rule is that it is entirely parameter- and model-free, and hence it is not subject to estimation risk given an ex ante allocation to the stock. Hence, it is not surprising that the optimal GMA rule discussed below is robust to parameter uncertainty and does not require any prior estimate of the predictive parameter. In contrast, the performances of the optimal dynamic rules depend on the accuracy of the estimates of the true parameters, which in turn depends not only on the sample size, but also on the prior.

In a continuous-time model, under fairly general conditions, one can separate the estimation from the optimization problem (see, e.g., Gennotte, 1986), and parameter uncertainty affects the optimal portfolio choice through dynamic learning. Barberis (2000) and Xia (2001), among others, show that this dynamic learning effect changes the myopic portfolio holding and adds a new component to dynamic hedging arising from the parameter uncertainty. For tractability, we follow the Xia approach to model uncertainty about predictability to examine the usefulness of the GMA rule. In this case, the stock price dynamics can be re-parameterized as

$$\frac{dS_t}{S_t} = (\mu_0 + \mu_1\bar{X} + \beta\hat{X}_t)\,dt + \sigma_s\,dB_t, \tag{67}$$

$$dX_t = (\theta_0 + \theta_1 X_t)\,dt + \sigma_x\,dZ_t, \tag{68}$$

where β is an unknown parameter to be inferred from the data. Uncertainty associated with β obviously measures an investor's uncertainty about predictability. All other parameters are assumed known. In particular, the long-term mean stock return, $\mu_0 + \mu_1\bar{X}$, is known, where $\bar{X} = -\theta_0/\theta_1$ is the long-term mean of X_t. Assume β follows a diffusion process

$$d\beta = \lambda(\bar{\beta} - \beta)\,dt + \sigma_\beta\,dZ_t^\beta, \tag{69}$$

where the parameters of this process, i.e., the long-term mean $\bar{\beta}$ and reversion speed λ, are known to investors. But the investor does not observe the innovation process Z_t^β directly and has to infer the realization of β through observations on S_t and X_t. To complete the model, assume $E(dB_t\,dZ_t^\beta) = \rho_{\beta s}\,dt$, $E(dZ_t\,dZ_t^\beta) = \rho_{\beta x}\,dt$, and $E(dB_t\,dZ_t) = \rho\,dt$.

Let $\mathscr{F}_t$ be the investor's filtration. Adapted to $\mathscr{F}_t$, the least square estimate of β is Gaussian, with mean and variance

$$b_t = E[\beta_t|\mathscr{F}_t] \tag{70}$$

and

$$v_t = E[(\beta_t - b_t)^2|\mathscr{F}_t]. \tag{71}$$

Starting from a Gaussian prior for β with mean b_0 and variance v_0, the Bayesian updating rule for the conditional mean and variance, b_t and v_t, are (see, Xia, 2001)

$$db_t = \lambda(\bar{b} - b_t)\,dt + \nu_1\,d\hat{B}_t + \nu_2\,d\hat{Z}_t, \tag{72}$$

$$\frac{dv_t}{dt} = -2\lambda v_t + \sigma_\beta^2 - (\nu_1^2 + \nu_2^2 + 2\nu_1\nu_2\rho), \tag{73}$$

where

$$\bar{b} = \bar{\beta}, \tag{74}$$

$$\nu_1 = \frac{v_t(X_t - \bar{X}) + \sigma_s\sigma_\beta(\rho_{\beta s} - \rho_{\beta x}\rho)}{\sigma_s(1-\rho^2)}, \tag{75}$$

$$\nu_2 = \frac{-v_t(X_t - \bar{X})\rho_{xs} + \sigma_s\sigma_\beta(\rho_{\beta x} - \rho_{\beta s}\rho)}{\sigma_s(1-\rho^2)}, \tag{76}$$

$$d\hat{B}_t = dB_t + \frac{(X_t - \bar{X})(\beta_t - b_t)}{\sigma_s}\,dt, \tag{77}$$

and

$$d\hat{Z}_t = dZ_t. \tag{78}$$

To further simplify the problem, we assume log-utility. In this case, the optimal dynamic stock allocation can be solved analytically,

$$\xi_{\mathrm{opt}}^* = \frac{\mu_s + b_t(X_t - \bar{X}) - r}{\sigma_s^2}. \tag{79}$$

Hence, the optimal log-utility level is

$$\begin{aligned} U_{\mathrm{opt}}^* &= E\log W_T \\ &= \int_0^T E\Big[r + \xi_{\mathrm{opt}}^*(\mu_0 + \mu_1\bar{X} + \beta(X_t - \bar{X}) - r) \\ &\quad - \frac{1}{2}\xi_{\mathrm{opt}}^{*2}\sigma_s^2\Big]\,dt + \log W_0. \end{aligned} \tag{80}$$

This value function can be computed easily via simulation.

In particular, the optimal fixed rule in the parameter uncertainty case, under the log-utility, can be explicitly obtained as

$$\xi_{\mathrm{fix}}^* = \frac{\mu_s - r + C_T}{\sigma_s^2}, \tag{81}$$

where

$$C_T = \frac{1}{T}\int_0^T E[\beta\hat{X}_t]\,dt = \frac{\rho_{\beta x}\sigma_\beta\sigma_x}{(\theta_1 - \lambda)^2}\left[\frac{e^{(\theta_1-\lambda)T} - 1}{T} - 1\right]. \tag{82}$$

Intuitively, C_T captures the covariance between the predictability parameter β and state variable X_t.

For applications later, we summarize the three strategies in our parameter uncertainty setting.

1. The optimal dynamic learning rule ξ_{opt}^* as given by Eq. (79).
2. The optimal fixed strategy ξ_{fix}^* as given by Eq. (81).

3. The GMA rule, a combination of ξ^*_{fix} and the MA, with coefficients

$$\xi_{\text{fix}} = \xi^*_{\text{fix}} - \frac{\bar{\beta} b_1}{b_2(1-b_2)\sigma_s^2} \tag{83}$$

and

$$\xi_{\text{mv}} = \frac{\bar{\beta} b_1}{b_2(1-b_2)\sigma_s^2}, \tag{84}$$

where b_1 and b_2 are defined similarly in Eqs. (27) and (28) with the unknown μ_1 now replaced by the long-term mean $\bar{\beta}$.

The fixed and GMA rules are denoted as Fix1 and Fix1 + MA, respectively, because they are the corresponding strategies of the complete information case.

3.5. Solutions under model uncertainty

In this subsection, we consider further the case in which the true model is not completely known to investors. Previously, knowledgable investors could obtain their optimal trading strategies based on their assumed true model, but now the true model is unknown both to these smart investors and to the technical traders. To examine how well the GMA strategy performs in this seemingly realistic case, given that no one knows the exact model of stock prices, we need first to provide a way for constructing the optimal GMA. We have solved the optimal GMA strategy in terms of the true parameters of the model, but this is not absolutely necessary. We show now that the optimal GMA strategy can be estimated with much less model dependence. In other words, the strategy is robust to a wide class of model specifications. To see this, assume now that we have a very general stock price process

$$\frac{dS_t}{S_t} = R_t\,dt + \sigma\,dB_t, \tag{85}$$

where R_t is the instantaneous expected stock return that can be stochastic. For simplicity, σ is assumed, as before, as the constant volatility parameter. Then the log wealth process of the GMA strategy is

$$\begin{aligned}\log W_T = {} & \log W_0 + rT + \int_0^T (\xi_{\text{fix}} + \xi_{\text{mv}}\eta_t)(R_t - r)\,dt \\ & + \int_0^T (\xi_{\text{fix}} + \xi_{\text{mv}}\eta_t)\sigma\,dB_t \\ & - \frac{1}{2}\int_0^T (\xi_{\text{fix}} + \xi_{\text{mv}}\eta_t)^2\sigma^2\,dt.\end{aligned} \tag{86}$$

Hence, the expected utility becomes

$$\begin{aligned}U &= \mathrm{E}\log W_T \\ &= \log W_0 + rT + (\xi_{\text{fix}} b_0 + \xi \text{mv} b_1 - \tfrac{1}{2}\xi_{\text{fix}}^2\sigma^2 \\ &\quad - \xi_{\text{fix}}\xi_{\text{mv}}\sigma^2 b_2 - \tfrac{1}{2}\xi_{\text{mv}}^2\sigma^2 b_2)T,\end{aligned} \tag{87}$$

where

$$b_0 = \frac{1}{T}\int_0^T \mathrm{E}[R_t - r]\,dt, \tag{88}$$

$$b_1 = \frac{1}{T}\int_0^T \mathrm{E}[\eta_t(R_t - r)]\,dt, \tag{89}$$

and

$$b_2 = \frac{1}{T}\int_0^T \mathrm{E}\eta_t\,dt. \tag{90}$$

Optimizing the expected utility, we obtain

$$\hat{\xi}^*_{\text{fix}} = \frac{b_0}{\sigma^2} - b_2\hat{\xi}^*_{\text{mv}} \tag{91}$$

and

$$\hat{\xi}^*_{\text{mv}} = \frac{1}{\sigma^2(1-b_2)}\left(\frac{b_1}{b_2} - b_0\right). \tag{92}$$

The parameters defined in Eqs. (88)–(90) can be written in terms of moments,

$$b_0 = \mathrm{E}[R_t] - r, \tag{93}$$

$$b_1 = \mathrm{E}[\eta_t R_t] - rb_2, \tag{94}$$

and

$$b_2 = \mathrm{E}[\eta_t]. \tag{95}$$

Thus, assuming stationarity as before, we can estimate them by their sample analogues. For example, to see how b_1 can be estimated, we write

$$R_t\Delta t = \frac{\Delta S_t}{S_t} - \sigma\Delta B_t. \tag{96}$$

With the law of iterative expectation, we have

$$b_1 = \mathrm{E}[\eta_t \mathrm{E}_t(R_t - r)] = \mathrm{E}\left[\eta_t\left(\frac{\Delta S_t}{S_t\Delta t} - r\right)\right], \tag{97}$$

which can be estimated by using the corresponding sample average of the right-hand side.

Now we are ready to define the estimated optimal GMA strategy as follows (which differs from the optimal GMA that solves from a given specification of the true model). At any time t, we use the available sample moments up to that time to estimate the parameters given by Eqs. (93)–(95). Substituting the estimates into Eqs. (91) and (92), we obtain the estimated optimal GMA strategy $\hat{\xi}^*_{\text{fix}} + \hat{\xi}^*_{\text{mv}}\eta_t$. Because the estimates $\hat{\xi}^*_{\text{fix}}$ and $\hat{\xi}^*_{\text{mv}}$ vary over time according to the moment estimates at time t and do not depend on future information, the strategy is a feasible rolling strategy. No knowledge of the true model is needed other than the general form of Eq. (85). The GMA strategy, denoted later as Fix1 + MA, is robust to model specifications and outperforms the optimal trading strategies substantially when they are derived from the wrong models.

3.6. Optimal lags

So far, we have studied the various GMA strategies with a fixed lag. In this subsection, we ask how the lag can be optimized. We study this problem under the log-utility with the aid of the analytical solutions of Section 3.2. However, the optimal lag itself does not admit an explicit solution but can be solved approximately in closed form that provides qualitative insights on the driving factors.

Unlike the previous two subsections, we assume here as usual that the investor knows all the true parameters of the model to simplify the analysis.

To study the determinants of the optimal lag, we restrict parameter values to those of practical interest by assuming

$$\sigma_s^2 \gg \frac{\mu_1^2\sigma_x^2}{\theta_1^2} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}. \tag{98}$$

This is because σ_x is much smaller relative to σ_s, and because the correlation ρ is close to zero. This relation holds for all three calibrated models provided later. Using the unit-free variable $x = \sqrt{|\theta_1|L}$, we can approximate Eqs. (29)–(31) by

$$C_{12}^Z \approx C_1\left(1 - \frac{1-e^{-x^2}}{x^2}\right), \tag{99}$$

$$C_{22}^Z \approx \frac{\sigma_s^2}{3}L = C_2x^2, \tag{100}$$

$$m_2^Z = \frac{\mu_s - \sigma_s^2/2}{2}L = C_3x^2, \tag{101}$$

where

$$C_1 = \frac{\mu_1\sigma_x^2}{2\theta_1^2} - \frac{\sigma_x\sigma_s\rho}{\theta_1}, \tag{102}$$

$$C_2 = \frac{\sigma_s^2}{3|\theta_1|}, \tag{103}$$

and

$$C_3 = \frac{\mu_s - \sigma_s^2/2}{2|\theta_1|}. \tag{104}$$

Therefore, Eqs. (27) and (28) can be approximated by

$$b_1 \approx C_4 \cdot \frac{1}{x}\left(1 - \frac{1-e^{-x^2}}{x^2}\right) \cdot f(Ax) = C_4h(x)f(Ax), \tag{105}$$

$$b_2 \approx \mathrm{N}(Ax), \tag{106}$$

where

$$A = \frac{C_3}{\sqrt{C_2}} = \frac{\sqrt{3}}{2} \cdot \frac{\mu_s - \frac{\sigma_s^2}{2}}{\sigma_s\sqrt{|\theta_1|}}, \tag{107}$$

$$C_4 = \frac{C_1}{\sqrt{C_2}}, \tag{108}$$

$$h(x) = \frac{1}{x}\left(1 - \frac{1-e^{-x^2}}{x^2}\right), \tag{109}$$

and $f(\cdot)$ is the standard normal density function. Then, we are ready to present Proposition 4.

Proposition 4. *In the class of strategies* GMA(S_t, G_t, γ)*, if the investment horizon T is long enough, then the optimal lag L_{opt} under the log-utility is approximately given by*

$$L_{\text{opt}} \approx \left[|\theta_1|\left(\frac{1+A_i^2}{2} + \sqrt{\left(\frac{1+A_i^2}{2}\right)^2 - \left(\frac{5}{12}+\frac{A_i^2}{3}\right)^2}\right)\right]^{-1}, \tag{110}$$

where $A_i = A/\sqrt{2}$ and A for the PureMA and Fix1 + MA *strategies, respectively.*

Proposition 4 says that optimal lag is mainly a function of the unconditional mean return μ_s, stock volatility σ_s, and state variable mean reversion speed $|\theta_1|$ given that T is large. Because μ_s and σ_s are stable across different models, L_{opt} is mainly driven by differences in θ_1.

Finally, consider the optimal lag for the pure MA strategy. Intuitively, given a lag length, the initial value of the MA matters little when T is large. However, given T, the initial value matters significantly in choosing L. This is because L can be chosen as T. Because the pure MA underperforms ξ_{fix1}^* under the practical parameter values, it is optimal to let $L = T$. In this case, the pure MA is identical to Fix1 because the initial value is chosen as ξ_{fix1}^*. An alternative initial value for the pure MA is zero. In this case, it can be shown that

$$L_{\text{opt}} \approx \frac{2\log(|\theta_1|T)}{A|\theta_1|} \tag{111}$$

when $|\theta_1|T$ is large. This makes intuitive sense. The larger the speed of mean reversion, the shorter the lag length to capture the change of trends.

4. An empirical illustration

To get further insights into the practical importance of technical analysis, we, in this section, calibrate the model from real data and compare the performance of various trading strategies in three cases. In the first case, with power-utility and with complete information, we examine the performance of the two fixed strategies and their combinations with the MA, Fix1, Fix2, Fix1 + MA, Fix2 + MA, and PureMA, relative to the performance of the dynamic optimal strategy. To make the comparison comprehensive, we also include three ad hoc MA strategies, MA1, MA2, and MA3, with stock allocations of 100%, Fix1 and Fix2, respectively, when the MA indicates a buy signal, and nothing otherwise. In addition, we consider the linear strategy of Aït-Sahalia and Brandt (2001) and Brandt and Santa-Clara (2006).[10] In the second case, under parameter uncertainty, we consider the log-utility and examine the relative performance of Fix1 and Fix1 + MA only. This is because Fix2 reduces to Fix1 and Fix2+MA reduces to Fix1 + MA, and because the remaining strategies, the ad hoc MAs and the linear, do not perform well and hence are omitted. In the third case, under model uncertainty, we examine only the estimated Fix1 and Fix1 + MA because they are unknown and have to be estimated from available realizations. For clarity, Table 1 summarizes the cases and the strategies used in the comparisons.

The data used in the calibration below are the monthly returns from December 1926 to December 2004 on Standard and Poor's 500 and monthly observations on three popular variables, the dividend yield, term-spread and payout ratio, which are used, respectively, as the

[10] See Section A.6 of Appendix A for more discussion and for the implementation details.

Table 1
List of various portfolio strategies and their comparisons.
The table lists all the strategies to be compared with the optimal dynamic strategy in three cases for the predictive model of the stock price: complete information, parameter uncertainty, and model uncertainty. There are nine strategies in the first case and two strategies in other two cases. The strategy Fix1 is the standard fixed allocation rule that invests a fixed proportion of wealth, determined by the unconditional moments of the model, into the stock, and Fix2 is also a fixed rule but accounting for stock predictability. The strategies Fix1 + MA and Fix2 + MA are those that are optimally combined with the MA. PureMA is the strategy that uses the MA optimally to time the stock without any combination with the fixed rules. MA1, MA2, and MA3 are ad hoc MA only strategies whose stock allocations are 100%, Fix1, and Fix2, respectively, when the MA indicates a buy signal (i.e., current stock price is above MA) and nothing otherwise. The final strategy, the linear rule, is the approximate linear portfolio policy of Brandt and Santa-Clara (2006).

Case 1 (complete information)	Case 2 (parameter uncertainty)	Case 3 (model uncertainty)
Fix1	Fix1	Fix1
Fix2		
Fix1 + MA	Fix1 + MA	Fix1 + MA
Fix2 + MA		
PureMA		
MA1		
MA2		
MA3		
Linear rule		

predictive variable in the model.[11] With the calibrated model and with setting $\gamma = 2$ and $r = 5\%$, we are ready to compute all the quantities of interest via simulations based on our analytical results in Section 3. We report below primarily the certainty-equivalent (CE) losses of the strategies as compared with the optimal dynamic one, which are easier to interpret than the utility values.

The CE losses are computed as follows. Normalizing the initial wealth at the level of \$100, $W_0 = 100$. Let $U^*_{\text{opt}}(W_0)$ be the expected utility based on the optimal dynamic strategy and $U^*_f(W_0)$ be the expected utility based on any of the suboptimal trading strategies, say a fixed strategy. Because $U^*_{\text{opt}}(W_0) \geqslant U^*_f(W_0)$, there exists $\text{CE} \geqslant 0$ such that

$$U^*_{\text{opt}}(W_0 - \text{CE}) = U^*_f(W_0). \tag{112}$$

CE can be interpreted as the perceived certainty-equivalent loss at time zero to an investor who switches the optimal strategy to the suboptimal one. In other words, the investor would be willing to give up *CE* percent of his initial wealth to avoid investing in the suboptimal strategy. Similar measures are used by Kandel and Stambaugh (1996), Pástor and Stambaugh (2000), Fleming, Kirby, and Ostdiek (2001), and Tu and Zhou (2004), among others. For simplicity, we refer to the CE as utility gains or losses in what follows.

[11] See, e.g., Goyal and Welch (2003) for a detailed description of the predictive variables, which are available from Goyal's website.

Table 2
Calibrated model parameters.
The table reports parameter estimates for the following cum-dividend price process:

$$\frac{dS_t}{S_t} = (\mu_0 + \mu_1 X_t)\,dt + \sigma_s\,dB_t,$$
$$dX_t = (\theta_0 + \theta_1 X_t)\,dt + \sigma_x\,dZ_t,$$

where $\mu_0, \mu_1, \sigma_s, \theta_0, \theta_1$, and σ_x are parameters, X_t is a predictive variable, and B_t and Z_t are standard Brownian motions with correlation coefficient ρ. The estimation is based on monthly returns on the Standard and Poor 500 from December 1926 to December 2004 and on X_t, which is the dividend yield, term-spread, and payout ratio, respectively, in the corresponding time period.

Parameters	Dividend yield	Term-spread	Payout ratio
μ_0	0.031	0.097	0.282
μ_1	2.072	1.206	−0.292
σ_s	0.195	0.195	0.194
θ_0	0.010	0.009	0.014
θ_1	−0.253	−0.527	−0.027
σ_x	0.012	0.013	0.050
ρ	−0.073	0.001	−0.003

4.1. Comparison under complete information

For the empirical results, we first report in Table 2 the calibrated parameters (whose estimation details are provided in Section A.5 of Appendix A). As expected, the stock volatility estimates are virtually the same as $\sigma_s = 0.195$ across the three predictive models. The same is true for the long-term mean of the stock return (not shown in the table). However, both the volatility of the predictive variable and its correlation with the stock return do vary across the models, making the comparison of the strategies more interesting.

Tables 3 and 4 report the CE losses in percentage points when $L = 50$ and 200 days, respectively.[12] The lag lengths are those used by Brock, Lakonishok, and LeBaron (1992), of which $L = 200$ is also the lag length of the popular MA chart published by *Investor's Business Daily*, the major competitor of the *Wall Street Journal*. There are several interesting facts. First, the losses are substantial across all the strategies relative to the optimal dynamic one, and they vary substantially, too, across predictive models. When the predictive variable is taken as the dividend yield, the losses (ignoring the ad hoc MA and linear strategies, which are dropped later for reasons below) vary from 7.895% to 50.356%. The range widens, from 18.061% to 59.359%, when the payout ratio is taken as the predictive variable. However, it narrows down to a low of 1.550% and a high of 42.910% when the term-spread is taken as the predictive variable. The large losses suggest strongly that, in an asset allocation problem, it is very important to know both the true dynamics of stock returns and the associated optimal dynamic strategy. This could help explain why Wall Street firms spend enormous amounts of money collecting data and doing research. Kandel and Stambaugh (1996) show that the economic loss can be significant when one ignores predictability

[12] The results when $L = 100$ are similar and omitted for brevity.

Table 3
Utility losses versus optimal strategy ($L = 50$).

The table reports the utility losses, measured as percentage points of initial wealth, that one is willing to give up to switch from a given strategy to the optimal dynamic one in the complete information model when the moving average (MA) lag length L is set equal to 50 days.

Strategy	Dividend yield	Term-spread	Payout ratio
$T = 10$			
Fix1	8.845	3.895	20.856
Fix2	7.904	1.568	18.061
Fix1 + MA	8.177	2.615	18.639
Fix2 + MA	7.895	1.550	18.061
Pure MA	16.303	13.088	27.692
MA1	17.762	14.037	28.096
MA2	17.234	14.423	30.815
MA3	16.863	13.614	28.432
Linear rule	14.801	13.702	33.231
$T = 20$			
Fix1	16.680	7.609	31.275
Fix2	15.171	3.112	30.682
Fix1 + MA	15.344	4.648	30.619
Fix2 + MA	15.161	3.059	30.681
Pure MA	29.326	23.931	41.309
MA1	31.572	25.652	43.109
MA2	30.694	26.254	43.974
MA3	30.040	24.874	42.992
Linear rule	33.409	30.891	58.311
$T = 40$			
Fix1	30.369	14.613	50.694
Fix2	28.027	6.129	49.495
Fix1 + MA	27.907	7.949	50.272
Fix2 + MA	27.985	5.880	49.495
Pure MA	50.356	42.910	59.359
MA1	53.684	45.284	63.632
MA2	51.972	45.561	60.039
MA3	51.136	43.758	61.304
Linear rule	61.541	58.711	80.391

Table 4
Utility losses versus optimal strategy ($L = 200$).

The table reports the utility losses, measured as percentage points of initial wealth, that one is willing to give up to switch from a given strategy to the optimal dynamic one in the complete information model when the moving average (MA) lag length L is set equal to 200 days.

Strategy	Dividend yield	Term-spread	Payout ratio
$T = 10$			
Fix1	8.845	3.895	20.856
Fix2	7.904	1.568	18.061
Fix1 + MA	8.125	2.497	18.145
Fix2 + MA	7.896	1.547	18.059
Pure MA	15.181	11.526	24.846
MA1	17.285	14.099	25.881
MA2	16.383	13.642	28.193
MA3	16.183	13.385	26.047
Linear rule	14.801	13.702	33.231
$T = 20$			
Fix1	16.680	7.609	31.275
Fix2	15.171	3.112	30.682
Fix1 + MA	14.992	4.460	30.521
Fix2 + MA	15.168	3.040	30.682
Pure MA	26.569	21.435	38.720
MA1	30.642	24.809	41.252
MA2	29.272	24.662	41.506
MA3	28.756	23.694	40.591
Linear rule	33.409	30.891	58.311
$T = 40$			
Fix1	30.369	14.613	50.694
Fix2	28.027	6.129	49.495
Fix1 + MA	27.341	6.975	50.694
Fix2 + MA	28.021	5.887	49.495
Pure MA	45.315	36.686	55.404
MA1	49.746	40.360	59.826
MA2	48.755	41.713	56.518
MA3	47.520	39.032	57.957
Linear rule	61.541	58.711	80.391

completely when a small degree of predictability exists in the data. In a continuous-time version of their model, this is apparent when we examine the losses of Fix1 versus the optimal dynamic strategy. However, the optimal dynamic strategy is difficult to identify, while the fixed rules are more practical and easy to apply. Even if the optimal dynamic rule is available, the predictive variable(s) might not be available at all time frequencies while the stock price can be observed virtually continuously during trading hours for implementing any MA-based strategies.

Second, Fix2 performs better than Fix1, which is not surprising because Fix1 is optimal only under the iid assumption. The superior performance varies across predictive variables and achieves the best level when the term-spread is taken as the predictive variable. The performance difference is of significant economic importance even when $T = 10$. This suggests that ignoring predictability entirely can lead to substantial economic losses even within the class of fixed strategies.

Third, the MA rule adds value to both Fix1 and Fix2, and Fix2+MA is the best suboptimal strategy. For Fix1, the MA improves its performance substantially by cutting the losses by at least 1–2% as long as $T > 10$. However, the MA provides only small improvement over Fix2. This does not suggest necessarily that the practical value of the MA rule is small. In practice, it is extremely difficult to know what process the stock follows and what variables drive the market from time to time. However, the long-term stock return and volatility could be estimated with little error due to the long historical data. This means that Fix1 is a feasible strategy while Fix2 might not be, at least to a sizable number of investors. By the same token, the dynamic optimal rule is difficult to identify in practice. Currently, index funds hold about one-third of the stocks. Such investors are likely to invest their money with allocations that resemble Fix1, not Fix2. In addition, popular portfolio optimization strategies (see, e.g., Litterman, 2003; Meucci, 2005) are more like Fix1 than Fix2. To the extent that this is true, the MA rule can have value. Theoretically, uncertainty about the degree of predictability can make the MA rule add value to the optimal dynamic rule, too, when the prior is not informative enough. There might be countless other reasons for the usefulness of the MA rule because so many successful practitioners put their money behind it in reality.

Fourth, the lag length makes only a small difference in the results except for the pure MA rule (and the ad hoc ones), which by definition depends on L more heavily. Because the fixed rules are independent of L, their values are the same across Tables 3 and 4. For both Fix1 + MA and Fix2+MA, their values change only from 8.177% and 7.895% to 8.125% and 7.896%, respectively, in the dividend yield model with $T=10$. When $T=40$, the values are larger and so are the differences. But the larger differences are still less than 0.5%. In contrast, for the PureMA, the largest difference is as high as about 5%, occurring at $T=40$.

Fifth, PureMA rules are much worse than other rules (except the ad hoc MA ones). For example, when the dividend yield is taken as the predictive variable and $L=50$, it has a loss about twice as large as the fixed rules when $T=10$. The qualitative results change little as T increases. When the term-spread is taken as the predictive variable, the difference can be four times as large. The least difference, still over 5%, occurs when the payout ratio is taken as the predictive variable. The results suggest strongly that one should not use MA alone, but only use it in conjunction with the fixed strategies.

Sixth, the ad hoc MA rules (MA1, MA2, and MA3) perform worse than PureMA. Theoretically, this is expected because the latter is optimal among pure MA rules. However, what is of interest here is that the under-performance can be of significant economic importance. Because these ad hoc rules perform poorly and do not add much information in comparison with other rules once we keep PureMA, we eliminate them henceforth.

Seventh, the linear rule under-performs the fixed rules and hence also their combinations with the MA. However, it outperforms the PureMA as well as the ad hoc MAs when $T=10$, but it does poorly when $T=20$ and 40. The results are not surprising. As shown by Brandt and Santa-Clara (2006) in their Table 1, the linear rule works well with 1% errors when the investment horizon is two years or so, but the error can increase to the order of 10% when the horizon lengthens to 10 years. There are two reasons that this happens. First, the linear approximation worsens as T gets greater. Second, the fourth-order polynomial approximation to the power-utility becomes worse as the horizon lengthens. Similar to the case with the ad hoc MA rules, for brevity, we no longer report the linear rule in what follows.

Now, let us examine the impact of using either arithmetic MAs or the ex-dividend stock prices in the computation of various strategies. To see the influence of the first, Table 5 reports the same valuation as Table 4 except that it replaces the previous geometric MAs with the arithmetic ones. The results are little changed. For example, when $T=40$ and when the dividend yield is taken as the predictive variable, Fix1 + MA has a value of 27.378%, which is virtually identical to the earlier value of 27.341%. The largest difference occurs for PureMA, which is still less than 0.5%. To see the effects of dividends, Table 6 computes the losses of Table 4 by using the ex-dividend prices instead, with an assumed annual dividend yield of 3%. Although the differences are larger now, they are confined only to PureMA. They make no difference whatsoever for other GMA strategies. Overall, we find that our earlier conclusions are robust to using either arithmetic averages or ex-dividend stock prices in the implementation of the fixed rules and their combinations with the MA.

Table 5
Utility losses versus optimal strategy for arithmetic average.
The table reports the utility losses, measured as percentage points of initial wealth, that one is willing to give up to switch from a given strategy to the optimal dynamic one in the complete information model when the moving average (MA) lag length L is set equal to 200 days, and when it is computed based on the arithmetic average instead of the geometric average.

Strategy	Dividend yield	Term-spread	Payout ratio
$T=10$			
Fix1	8.845	3.895	20.856
Fix2	7.904	1.568	18.061
Fix1 + MA	8.091	2.514	18.153
Fix2 + MA	7.901	1.547	18.070
PureMA	15.074	11.668	24.951
$T=20$			
Fix1	16.680	7.609	31.275
Fix2	15.171	3.112	30.682
Fix1 + MA	15.036	4.391	30.531
Fix2 + MA	15.166	3.053	30.719
PureMA	26.873	21.355	38.917
$T=40$			
Fix1	30.369	14.613	50.694
Fix2	28.027	6.129	49.495
Fix1 + MA	27.378	6.697	50.660
Fix2 + MA	28.017	5.963	49.641
PureMA	45.715	36.287	55.856

Table 6
Utility losses versus optimal strategy with ex-dividend price.
The table reports the utility losses, measured as percentage points of initial wealth, that one is willing to give up to switch from a given strategy to the optimal dynamic one in the complete information model when the moving average (MA) lag length L is set equal to 200 days, and when it is computed based on the ex-dividend price instead of the cum-dividend price.

Strategy	Dividend yield	Term-spread	Payout ratio
$T=10$			
Fix1	8.845	3.895	20.856
Fix2	7.904	1.568	18.061
Fix1 + MA	8.153	2.737	18.149
Fix2 + MA	7.898	1.521	18.064
PureMA	16.085	13.215	25.741
$T=20$			
Fix1	16.680	7.609	31.275
Fix2	15.171	3.112	30.682
Fix1 + MA	15.164	4.517	30.559
Fix2 + MA	15.162	3.058	30.728
PureMA	28.607	23.031	40.078
$T=40$			
Fix1	30.369	14.613	50.694
Fix2	28.027	6.129	49.495
Fix1 + MA	27.379	6.735	50.820
Fix2 + MA	28.028	6.023	49.683
PureMA	47.172	38.688	56.718

Table 7
Performance statistics for dividend model.
The table reports performance statistics for various strategies in the complete information model when the moving average (MA) lag length L is set equal to 200 and when the predictive variable is the dividend yield. The annualized mean is the annualized expected holding period return (HPR), the annualized SD is the standard deviation of the annualized HPR, the Sharpe ratio is defined as the annualized mean excess HPR divided by the annualized SD, and MaxDD is the maximum drawdown. Other variables are defined similarly with the rates computed based on continuous compounding.

Statistics	Fix1	Fix2	Fix1 + MA	Fix2 + MA	PureMA
$T = 10$					
Annualized mean	8.703	8.563	8.578	8.531	7.328
Annualized median	8.708	8.568	8.657	8.539	7.094
Annualized SD	5.386	5.081	5.102	5.010	4.229
Annualized Sharpe	0.688	0.701	0.701	0.705	0.551
Skewness	−0.008	−0.008	−0.057	−0.022	0.241
Kurtosis	2.950	2.950	2.967	2.956	3.063
Max DD	31.349	29.511	30.181	29.217	20.949
$T = 20$					
Annualized mean	8.793	8.611	8.652	8.573	7.442
Annualized median	8.768	8.588	8.630	8.566	7.282
Annualized SD	3.922	3.646	3.711	3.590	3.043
Annualized Sharpe	0.967	0.990	0.984	0.995	0.803
Skewness	−0.108	−0.108	−0.150	−0.121	0.183
Kurtosis	2.891	2.891	2.934	2.902	2.830
Max DD	37.359	34.640	36.046	34.288	24.130
$T = 40$					
Annualized mean	8.825	8.619	8.682	8.579	7.442
Annualized median	8.872	8.665	8.739	8.625	7.367
Annualized SD	2.889	2.663	2.737	2.622	2.161
Annualized Sharpe	1.324	1.359	1.345	1.365	1.130
Skewness	−0.016	−0.016	−0.046	−0.025	0.205
Kurtosis	3.009	3.009	2.999	3.006	3.161
Max DD	42.718	39.299	41.349	38.921	27.387

Table 8
Comparison under parameter uncertainty ($T = 10$).
The table reports both the utilities of the optimal learning, the standard fixed, Fix1, and its optimal combination with the moving average (MA) strategies, Fix1 + MA, and the associated certainty-equivalent losses, measured as percentage points of initial wealth, relative to the optimal learning strategy. The MA length is 200 days and investment horizon is T set equal to 10 years. The predictability parameter β is captured by a mean-reverting process starting from its long-term level $\bar{\beta}_0 = 2.0715$. The standard normal prior on β_0 has a prior mean b_0 and standard deviation $\sqrt{v_0}$.

$\sqrt{v_0}$	U_{opt}	U_{Fix1}	$U_{Fix1+MA}$	CE_{Fix1}	$CE_{Fix1+MA}$
$b_0 = 0$					
1	1.137	1.014	1.020	12.270	11.670
2	1.138	1.014	1.020	12.402	11.802
3	1.134	1.014	1.020	11.961	11.361
4	1.121	1.014	1.020	10.671	10.071
$b_0 = 4$					
1	1.147	1.014	1.020	13.271	12.671
2	1.149	1.014	1.020	13.459	12.859
3	1.145	1.014	1.020	13.071	12.471
4	1.131	1.014	1.020	11.690	11.090
$b_0 = 6$					
1	0.999	1.014	1.020	−1.549	−2.149
2	1.015	1.014	1.020	0.090	−0.510
3	1.030	1.014	1.020	1.511	0.911
4	1.035	1.014	1.020	2.050	1.450
$b_0 = 7$					
1	0.888	1.014	1.020	−12.640	−13.240
2	0.915	1.014	1.020	−9.929	−10.529
3	0.942	1.014	1.020	−7.199	−7.799
4	0.961	1.014	1.020	−5.379	−5.979

Finally, to understand better the strategies, it is of interest to examine their performance statistics, i.e., the annualized mean, median, standard error, and Sharpe ratio, as well as the skewness, kurtosis, and maximum drawdown (MaxDD). The annualized mean is the annualized expected holding period return (HPR), the annualized SD is the standard deviation of the annualized HPR, and the Sharpe ratio is defined as the annualized mean excess HPR divided by the annualized SD. Other variables are defined similarly with the rates computed based on continuous compounding. Table 7 reports the results when the dividend yield is used as the predictive variable. The returns on both Fix1 and Fix2 are generally greater than those of their MA combinations, but their standard deviations are larger, too. Consequently, the Sharpe ratios of the fixed rules are smaller than those of the latter. This is consistent with the results from utility maximization. As expected, the Sharpe ratios increase as the horizon lengthens. The skewness and kurtosis for both the fixed strategies and their combinations are small. In contrast, the PureMA has relatively higher values. The same pattern also holds for the kurtosis. The MaxDDs, the average MaxDDs over the simulated paths of the model, are substantial for all the strategies, though those for the PureMA are much smaller.[13] It seems that one has to be prepared for the big ups and downs in long-term investments. Nevertheless, both Fix1 + MA and Fix2 + MA have smaller drawdowns than their counterparts. Similar results, omitted for brevity, are also obtained when either the term-spread or payout ratio is used as the predictive variable.

4.2. Comparison under parameter uncertainty

As in Xia (2001), we assume $\rho_{\beta x}$ to be zero. Then, neither Fix1 nor Fix1 + MA depends on the unknown parameter β, and ξ^*_{fix} reduces to the optimal fixed rule ξ^*_{fix2}. In addition, for the mean-reverting process on β, we assume β_t starts from its calibrated long-term mean, $\bar{\beta}_0 = 2.072$, and set the reverting speed $\lambda = 0.115$ and the volatility $\sigma_\beta = 1.226$.

The results are provided in Table 8 with the dividend yield as the predictive variable, $L = 200$ days and $T = 10$ years. The first two columns are values for the prior mean

[13] The same magnitude of drawdowns also shows up in the standard geometric Brownian motion model without the predictive component of our model here. Magdon-Ismail, Atiya, Pratap, and Abu-Mostafa (2004) provide an analytical analysis of the MaxDD for a Brownian motion.

and standard error, the third to the fifth columns are the expected utilities associated with the optimal learning strategy, Fix1 and Fix1 + MA, respectively. The last two columns are the CE or utility losses (in percentage points) of the Fix1 and Fix1 + MA relative to the optimal learning one. Because $\rho_{\beta x} = 0$, the performances of both Fix1 and Fix1 + MA are independent of priors on β. The performance of the optimal updating rule depends on the prior. When the prior mean $b_0 = 0$, both Fix1 and Fix1 + MA under-perform the optimal learning rule substantially, with losses from 10.67% to 12.40% and 10.07% to 11.80%, respectively. Among the priors, $\sqrt{v_0} = 2$ is clearly the best one, and hence it is not surprising to see that the associated loss is the largest. While it is unclear ex ante whether or not $\sqrt{v_0} = 1$ is better than $\sqrt{v_0} = 3$, the former turns out to provide a higher expected utility for the optimal learning. The reason is that the model seems to penalize large prior means b_0 more than small ones relative to the true β_0. This is why the losses become greater when $\sqrt{v_0}$ further increases from 3. When the prior mean $b_0 = 4$, the results are similar qualitatively. However, when the prior $b_0 = 6$, which is not too informative about the true β_0, the optimal learning rule can now perform worse than either Fix1 or Fix1 + MA when $\sqrt{v_0} = 1$. When the prior mean moves further away at $b_0 = 7$, the losses increase substantially to over 10%. The optimal learning also depends on the investment horizon.

Table 9

Comparison under parameter uncertainty ($T = 5$).

The table reports both the utilities of the optimal learning, the standard fixed, Fix1, and its optimal combination with the moving average (MA) strategies, Fix1 + MA, and the associated certainty-equivalent losses, measured as percentage points of initial wealth, relative to the optimal learning strategy. The MA length is 200 days and investment horizon is T set to five years. The predictability parameter β is captured by a mean-reverting process starting from its long-term level $\bar{\beta}_0 = 2.0715$. The standard normal prior on β_0 has a prior mean b_0 and standard deviation $\sqrt{v_0}$.

$\sqrt{v_0}$	U_{opt}	U_{Fix1}	$U_{Fix1+MA}$	CE_{Fix1}	$CE_{Fix1+MA}$
$b_0 = 0$					
1	0.503	0.457	0.460	4.590	4.231
2	0.504	0.457	0.460	4.681	4.322
3	0.501	0.457	0.460	4.381	4.021
4	0.491	0.457	0.460	3.470	3.111
$b_0 = 4$					
1	0.514	0.457	0.460	5.770	5.411
2	0.515	0.457	0.460	5.799	5.440
3	0.511	0.457	0.460	5.399	5.040
4	0.500	0.457	0.460	4.349	3.990
$b_0 = 6$					
1	0.404	0.457	0.460	−5.301	−5.660
2	0.414	0.457	0.460	−4.240	−4.599
3	0.423	0.457	0.460	−3.409	−3.769
4	0.424	0.457	0.460	−3.259	−3.618
$b_0 = 7$					
1	0.319	0.457	0.460	−13.750	−14.109
2	0.338	0.457	0.460	−11.920	−12.279
3	0.355	0.457	0.460	−10.149	−10.508
4	0.366	0.457	0.460	−9.090	−9.450

Table 10

Comparison under model uncertainty.

The table reports the utility losses of the estimated Fix1 and Fix1 + MA relative to the optimal strategies derived from the three predictive models with the dividend yield, term-spread and payout ratio as the predictive variable, respectively. In each of the three panels, the model associated with the variable name of the panel is assumed to be the true model, while the other two are the wrong models. The moving average (MA) lag length L is 50 or 200 days, and the investment horizon T is set equal to five, 10 and 20 years, respectively.

Investment	Fix1 + MA		Fix1	Uncertain models	
Horizon	$L = 50$	$L = 200$		Wrong Model 1	Wrong Model 2
Panel A: Dividend yield					
$T = 5$	5.228	5.333	5.616	6.593	17.288
$T = 10$	13.558	13.359	13.989	15.439	38.945
$T = 20$	28.294	27.848	28.366	31.009	70.774
Panel B: Term-spread					
$T = 5$	1.361	1.420	1.833	6.593	9.868
$T = 10$	3.892	3.652	5.408	15.439	23.433
$T = 20$	8.735	8.360	11.264	31.023	50.346
Panel C: Payout ratio					
$T = 5$	3.372	3.642	4.190	17.288	9.868
$T = 10$	12.313	12.724	16.431	38.945	23.433
$T = 20$	34.936	35.467	40.112	70.774	50.337

As the horizon shortens, the optimal learning becomes worse, as expected, as shown in Table 9 with $T = 5$ years. Overall, to the extent that uncertainty about predictability is high and the prior is not very informative, the widely used fixed strategy appears viable as it can outperform the optimal learning one. However, the MA rule can always add value to this fixed rule. Therefore, the MA rule or technical analysis seems capable of capturing information on the market that is useful to investors.

4.3. Comparison under model uncertainty

To assess the effect of model uncertainty, we assume that the true stock price process is one of the three calibrated models, but this is unknown to the investors. There are three cases to consider, each of which corresponds to one of the three models as the true one, respectively.[14] In the first case in which the model with the dividend yield as the predictive variable is assumed the true data-generating process, Panel A of Table 10 reports the utility losses by using the estimated Fix1 + MA and the optimal trading strategies based on the wrong models, the second and third one, respectively.[15]

[14] Model uncertainty is a real issue in practice as witnessed by the recent collapse of the supposed smart traders, the top investment banks. Zhou and Zhu (2009) examine the empirical performance of various GMA rules over the past century, and find that, if an individual had followed the GMA rule, his portfolio would have avoided much of the financial crisis, and could be better still if using some more technical rules.

[15] Although not reported, the estimated Fix1 + MA differs only slightly from the true one. For example, in the first case, when $T = 5$ and $L = 50$, their difference is less than 0.5%.

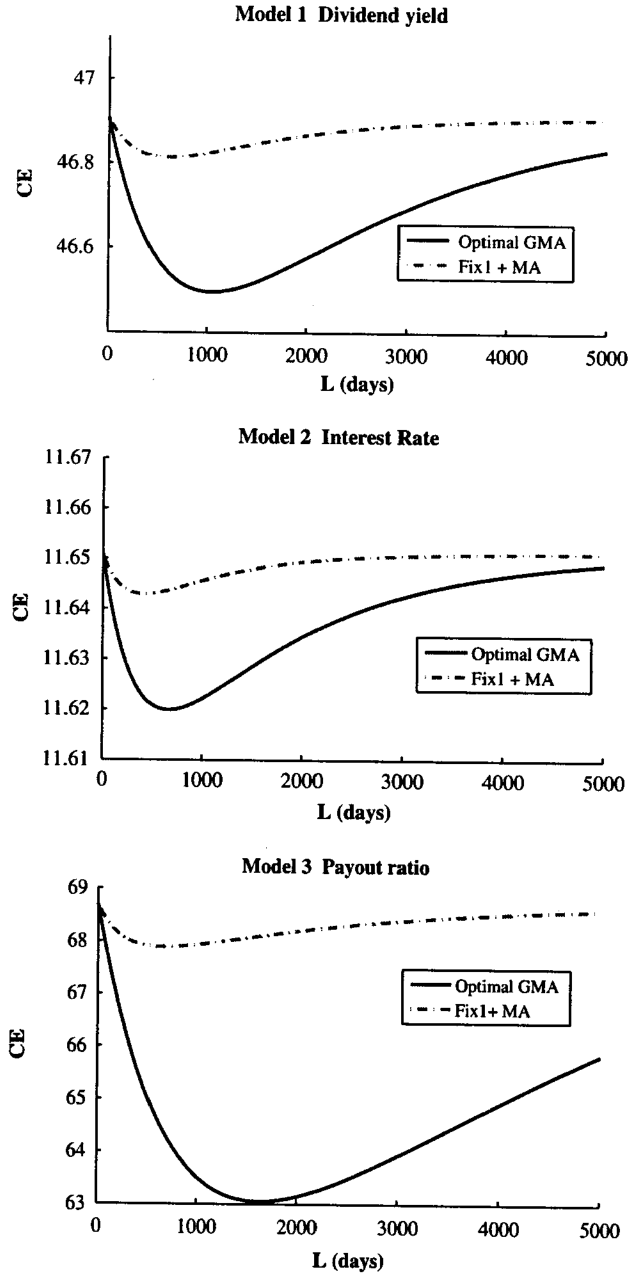

Fig. 1. Effect of lag length (L). The figure plots the certainty-equivalent (CE) losses versus the moving average lag length measured in days in the three predictive models. MA: moving average; GMA: generalized moving average.

As before, the losses here are measured relative to the true optimal strategy. When $T = 5$, the largest loss of Fix1 + MA is 5.333%, far smaller than 17.288%, the largest of the wrong optimal strategies. It is also smaller than 6.593%, the smallest of the latter. As investment horizon increases, the loss increases. The same conclusion also holds when the assumed true mode is either term-spread or payout ratio as the predictive variable, respectively, as indicated by the results in Panels B and C of the table.

An open question is how well Fix1 + MA compares with the estimated fixed strategy, i.e., $\hat{\xi}^*_{\text{fix}} = \hat{b}_0/\hat{\sigma}^2$ with $\hat{b}_0$ and $\hat{\sigma}^2$ as the moment estimators, which is denoted as Fix1. The utility losses associated with Fix1 are reported in the fourth column of Table 10. They are always larger than those associated with Fix1 + MA, and they are substantially so in many cases. This indicates that Fix1 + MA outperforms Fix1 not only when the true model is known, as is the case in Section 4.1, but also when the true model is unknown, as is the case here.

Overall, our results show that, while Fix1 + MA has lower utility than the true optimal one, it outperforms all the optimal strategies when they are derived from wrong models. Given that the true model is unknown and difficult to identify by investors in the real world, the robustness of Fix1 + MA, or of the technical analysis in general, makes it a valuable tool in practice.

4.4. The effect of lag lengths

The analytical optimal lags are available for both the optimal GMA and the Fix1 + MA strategy. Fig. 1 plots the utility losses of these two strategies relative to the optimal dynamic one at various lag lengths when $T = 40$. Because of differences in θ_1, as predicted by Proposition 4, the optimal lag in the term-spread model is the smallest and becomes the largest in the payout ratio model. There are in addition two interesting facts. First, the utility losses are much greater than those reported in Tables 3 and 4. This is expected because here $\gamma = 1$, while γ has a value of 2 in the tables. The smaller the γ, the more the risk taking, and so the greater the impact of the various stock allocation strategies on the expected utility. Second, the performance across different lags do not vary much for Fix1 + MA, implying that our earlier utility comparisons are insensitive to the use of the optimal lags. However, the optimal GMA rule is substantially more influenced by the use of the optimal lag than Fix1 + MA. But this does not affect our earlier results, because numerical studies on this rule are not provided due to the unavailability of its solution in the power-utility case.

5. Conclusion

Although technical analysis is popular in investment practice, few theoretical studies on it are available. The empirical evidence is mixed, and a lack of understanding exists on the economic rationale for its usefulness. In this paper, we provide a theoretical justification for an investor to use the MA rule, one of the widely used technical rules, in a standard asset allocation problem. The theoretical framework offers a number of useful insights about technical analysis. First, it solves the portion of investment a technical trader should allocate into the stock market if he receives a technical buy signal, while previous researchers determine it in ad hoc ways. Second, it shows how an investor might add value to his investment by using technical analysis, especially the MA, if he follows a fixed allocation rule that invests a fixed portion of wealth into the stock market, as dictated by the random walk theory of stock prices or by the popular mean–variance approach. In particular, our paper explains why both risk aversion and the degree of predictability (quality of signal) affect the optimal use of the MA. Third, when model parameters are unknown and have to be estimated from data, our asset allocation framework illustrates that the combination of the fixed rule with the MA can even outperform the optimal learning rule, which is prior dependent, when the prior is reasonable and yet not too informative. Finally, when the true model is unknown, as is the case in practice, we find that the optimal GMA is robust to model specification and outperforms the optimal dynamic strategies substantially when they are derived from the wrong models, suggesting that technical analysis provides useful information for asset allocation especially when we are uncertain above the driving force of the market.

For tractability, our exploratory study assumes a simple predictive process for a single risky asset and examines the simplest MA rule. Studies that allow for both more general processes (such as those with jumps, factor structures, and multiple assets) and more elaborate rules are clearly called for. Broadly speaking, asset pricing anomalies, such as the momentum effect, can also be regarded as profitable technical strategies that depend on historical price patterns. Questions remain open: What underlying asset processes permit such anomalies? What are the associated optimal investment strategies? Further issues to address are how past prices and trading volumes reveal the strategies of the major market players, with their incomplete and complementary information, and how their interactions determine asset prices. All of these are important and challenging topics for future research.

Appendix A

A.1. Proof of Eqs. (15), (27) and (28)

Let $y_t = \log S_t$. Then the model for the predictive variable and stock price process are

$$\begin{cases} dX_t = (\theta_0 + \theta_1 X_t)\,dt + \sigma_x\,dZ_t, \\ dy_t = (\mu_0 + \mu_1 X_t - \sigma_s^2/2)\,dt + \sigma_s\,dB_t, \end{cases} \tag{113}$$

where (Z_t, B_t) is a two-dimensional Brownian motion with correlation coefficient ρ.

To rule out any explosive behavior, we assume $\theta_1 < 0$ throughout, which is consistent with empirical applications. Furthermore, we assume that X_t is a stationary process for $t \geqslant 0$. Integrating Eq. (113) for X_t, we have

$$X_t = X_0 e^{\theta_1 t} - \frac{\theta_0}{\theta_1}(1 - e^{\theta_1 t}) + \sigma_x \int_0^t e^{\theta_1 (t-s)}\,dZ_s. \tag{114}$$

It follows that X_t is normally distributed with mean and covariance

$$\mathrm{E}X_t = \mathrm{E}X_0 e^{\theta_1 t} - \frac{\theta_0}{\theta_1}(1 - e^{\theta_1 t}) \tag{115}$$

and

$$\mathrm{cov}(X_t, X_s) = \left[V(0) - \frac{\sigma_x^2}{2\theta_1}(e^{-2\theta_1 t \wedge s} - 1)\right] e^{\theta_1 (t+s)}, \tag{116}$$

respectively, where $\mathrm{E}X_0$ and $V(0)$ are the mean and variance of X_0. Then, the steady state mean and variance of X_t can be obtained by taking $t \to +\infty$ in Eqs. (115) and (116), i.e.,

$$\bar{X} = -\frac{\theta_0}{\theta_1} \tag{117}$$

and

$$\bar{V}_x = -\frac{\sigma_x^2}{2\theta_1}. \tag{118}$$

The necessary and sufficient condition for X_t to be stationary for $t \geqslant 0$ is that X_0 start from the steady state, i.e., X_0 is normally distributed with mean $\bar{X}$ and variance $V(0)=\bar{V}_X$. Under the stationarity condition, the first two moments Eqs. (115) and (116) that characterize the distribution of X_t can thus be simplified as

$$\mathrm{E}X_t=\bar{X}=-\frac{\theta_0}{\theta_1} \tag{119}$$

and

$$\mathrm{cov}(X_t,X_s)=-\frac{\sigma_x^2}{2\theta_1}e^{\theta_1|t-s|}. \tag{120}$$

With initial conditions $X|_{t=0}=X_0$, $y|_{t=0}=y_0$, we integrate Eq. (113) to obtain

$$\begin{cases} X_t=X_0e^{\theta_1 t}-\dfrac{\theta_0}{\theta_1}(1-e^{\theta_1 t})+\sigma_x\displaystyle\int_0^t e^{\theta_1(t-s)}\,dZ_s,\\ y_t=y_0+\displaystyle\int_0^t(\mu_0+\mu_1X_s-\sigma_s^2/2)\,ds+\sigma_sB_t.\end{cases} \tag{121}$$

Let $M_t=\log G_t$, where G_t is the geometric MA at time t, then

$$M_t=\frac{1}{L}\int_{t-L}^t y_s\,ds. \tag{122}$$

To derive Eq. (15), we note, under constant holding ξ_{fix2}, the wealth process is

$$\begin{aligned}\log W_T=&\ \log W_0+rT+\xi_{\mathrm{fix2}}(\mu_0-r-\xi_{\mathrm{fix2}}\sigma_s^2/2)T\\ &+\xi_{\mathrm{fix2}}\mu_1\int_0^T X_t\,dt+\xi_{\mathrm{fix2}}\sigma_sB_T.\end{aligned} \tag{123}$$

Then, optimizing over ξ_{fix2} the power-utility

$$\begin{aligned}&\frac{1}{1-\gamma}\mathrm{E}[\exp((1-\gamma)\log W_T)]\\ &=\frac{1}{1-\gamma}\exp[(1-\gamma)(\log W_0+rT\\ &+\xi_{\mathrm{fix2}}(\mu_0-r-\xi_{\mathrm{fix2}}\sigma_s^2/2)T)]\\ &\times\mathrm{E}\exp\left[\left(\xi_{\mathrm{fix2}}\mu_1\int_0^T X_t\,dt+\xi_{\mathrm{fix2}}\sigma_sB_T\right)(1-\gamma)\right],\end{aligned} \tag{124}$$

we obtain the solution

$$\xi^*_{\mathrm{fix2}}=\frac{(\mu_0-r)+\mu_1\mathrm{E}\left[\frac{1}{T}\int_0^T X_t\,dt\right]}{\gamma\sigma_s^2-(1-\gamma)(\mu_1^2A+2\mu_1\sigma_sB)}, \tag{125}$$

where

$$A=\frac{1}{T}\mathrm{var}\left[\int_0^T X_t\,dt\right] \tag{126}$$

and

$$B=\frac{1}{T}\mathrm{cov}\left[\int_0^T X_t\,dt,B_T\right]. \tag{127}$$

With Eqs. (121) and (120), A and B can be simplified as

$$\begin{aligned}A&=\int_0^T dt\int_0^T ds\langle X_t X_s\rangle=-\frac{\sigma_x^2}{2\theta_1}\int_0^T dt\int_0^T ds\,e^{\theta_1|t-s|}\\ &=\frac{\sigma_x^2}{\theta_1^2}\left(T+\frac{1-e^{\theta_1T}}{\theta_1}\right)\end{aligned} \tag{128}$$

and

$$B=\int_0^T\langle X_t,B_T\rangle\,dt=\frac{\rho\sigma_x}{\theta_1}\left(\frac{e^{\theta_1T}-1}{\theta_1}-T\right), \tag{129}$$

where $\langle\cdot,\cdot\rangle$ denotes the covariance operator conditional on information at time 0 throughout the appendix for brevity, and we have made use of the following fact that for $t\leqslant T$

$$\begin{aligned}\langle X_t,B_T\rangle&=\sigma_x\int_0^t e^{\theta_1(t-s)}\langle dZ_s,B_T\rangle\\ &=\sigma_x\int_0^t\rho\,e^{\theta_1(t-s)}\,ds=\frac{\rho\sigma_x}{\theta_1}(e^{\theta_1t}-1).\end{aligned} \tag{130}$$

To derive Eqs. (27) and (28), taking expectation in Eq. (121) and making use of Eq. (120), we obtain

$$\mathrm{E}y_t=y_0+(\mu_0+\mu_1\bar{X}-\sigma_s^2/2)t \tag{131}$$

and

$$\mathrm{E}M_t=y_0+(\mu_0+\mu_1\bar{X}-\sigma_s^2/2)\left(t-\frac{L}{2}\right) \tag{132}$$

when $t>L$. These results allow us to compute the following second moments for $t>L$:

$$\langle X_t,X_{t-L}\rangle=-\frac{\sigma_x^2}{2\theta_1}e^{\theta_1L}, \tag{133}$$

$$\begin{aligned}\langle y_t,X_{t-L}\rangle&=\int_0^t\mu_1\langle X_s,X_{t-L}\rangle\,ds+\sigma_x\sigma_s\int_0^{t-L}e^{\theta_1(t-L-s)}\langle dW_s,B_t\rangle\\ &=\int_0^{t-L}\mu_1\langle X_s,X_{t-L}\rangle\,ds+\int_{t-L}^t\mu_1\langle X_s,X_{t-L}\rangle\,ds\\ &\quad+\sigma_x\sigma_s\rho\int_0^{t-L}e^{\theta_1(t-L-s)}\,ds\\ &=\frac{\mu_1\sigma_x^2}{2\theta_1^2}(2-e^{\theta_1(t-L)}-e^{\theta_1L})-\frac{\sigma_x\sigma_s\rho}{\theta_1}(1-e^{\theta_1(t-L)}),\end{aligned} \tag{134}$$

$$\begin{aligned}\langle X_t,y_{t-L}\rangle&=\int_0^{t-L}\mu_1\langle X_s,X_t\rangle\,ds+\sigma_x\sigma_s\int_0^t e^{\theta_1(t-s)}\langle dW_s,B_{t-L}\rangle\\ &=\left(\frac{\mu_1\sigma_x^2}{2\theta_1^2}-\frac{\sigma_x\sigma_s\rho}{\theta_1}\right)(e^{\theta_1L}-e^{\theta_1t}),\end{aligned} \tag{135}$$

and

$$\begin{aligned}\langle y_t,y_t\rangle&=\sigma_s^2t+\int_0^t\int_0^t\mu_1^2\langle X_s,X_u\rangle\,ds\,du+2\sigma_s\int_0^t\mu_1\langle X_s,B_t\rangle\,ds\\ &=\left(\sigma_s^2+\frac{(\mu_1\sigma_x)^2}{\theta_1^2}-\frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}\right)t\\ &\quad+\left(\frac{(\mu_1\sigma_x)^2}{\theta_1^3}-\frac{2\mu_1\sigma_s\sigma_x\rho}{\theta_1^2}\right)(1-e^{\theta_1t}),\end{aligned} \tag{136}$$

where we have used the fact $\langle X_s,B_t\rangle=\sigma_x\int_0^s e^{\theta_1(s-u)}\rho\,du$, for $s\leqslant t$, an equality

$$\int_0^t\int_0^t\langle X_s,X_u\rangle\,ds\,du=\frac{\sigma_x^2}{\theta_1^2}t+\frac{\sigma_x^2}{\theta_1^3}(1-e^{\theta_1t}) \tag{137}$$

and another equality

$$\begin{aligned}\langle y_t, y_{t-L}\rangle &= \langle y_{t-L}, y_{t-L}\rangle + \int_{t-L}^{t} \mu_1 \langle X_s, y_{t-L}\rangle\, ds \\ &= \left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_s\rho\sigma_x}{\theta_1}\right)(t-L) \\ &\quad + \left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_s\rho\sigma_x}{\theta_1^2}\right) \\ &\quad \times (1 - e^{\theta_1(t-L)} + e^{\theta_1 L} - e^{\theta_1 t}). \end{aligned} \tag{138}$$

Next, we compute the following second moments involving M_t using Eqs. (134) and (138):

$$\begin{aligned}\langle X_t, M_t\rangle &= \frac{1}{L}\int_{t-L}^{t} \langle y_s, X_t\rangle\, ds \\ &= \frac{1}{L}\left(-\frac{\mu_1\sigma_x^2}{2\theta_1^3} + \frac{\sigma_x\sigma_s\rho}{\theta_1^2}\right)(1 - e^{\theta_1 L}) \\ &\quad - \left(\frac{\mu_1\sigma_x^2}{2\theta_1^2} - \frac{\sigma_x\sigma_s\rho}{\theta_1}\right)e^{\theta_1 t}\end{aligned} \tag{139}$$

and

$$\begin{aligned}\langle y_t, M_t\rangle &= \frac{1}{L}\int_{t-L}^{t} \langle y_t, y_s\rangle\, ds \\ &= \left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_s\sigma_x\rho}{\theta_1}\right)\left(T - \frac{L}{2}\right) \\ &\quad + \left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_x\rho\sigma_s}{\theta_1^2}\right)(1 - e^{\theta_1 T}) \\ &\quad - \left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_x\rho\sigma_s}{\theta_1^2}\right)\frac{1}{\theta_1 L}(1 - e^{\theta_1 L} \\ &\quad - e^{\theta_1(T-L)} + e^{\theta_1 T}).\end{aligned} \tag{140}$$

Finally, to compute $\langle M_t, M_t\rangle$, we note first

$$\begin{aligned}M_t &= \frac{1}{L}\int_{t-L}^{t} y_s\, ds = \frac{1}{L}\int_0^L [y_{t-L} + (y_{t-L+s} - y_{t-L})]\, ds \\ &= y_{t-L} + \frac{1}{L}\int_0^L \hat{y}_{t-L+s}\, ds,\end{aligned} \tag{141}$$

where $\hat{y}_{t-L+s} = y_{t-L+s} - y_{t-L}$. Then, we can write $\langle M_t, M_t\rangle$ as

$$\begin{aligned}\langle M_t, M_t\rangle &= \left\langle \left(y_{t-L} + \frac{1}{L}\int_0^L \hat{y}_{t-L+s}\, ds\right), \left(y_{t-L} + \frac{1}{L}\int_0^L \hat{y}_{t-L+s}\, ds\right)\right\rangle \\ &= \langle \hat{M}_L, \hat{M}_L\rangle + \frac{2}{L}\int_0^L \langle y_{t-L}, y_{t-L+s}\rangle\, ds - \langle y_{t-L}, y_{t-L}\rangle,\end{aligned} \tag{142}$$

where $\hat{M}_t = (1/t)\int_0^t y_s\, ds$. Using Eq. (138), we obtain

$$\begin{aligned}\langle \hat{M}_t, \hat{M}_t\rangle &= \frac{1}{t^2}\int_0^t\int_0^t \langle y_s, y_u\rangle\, ds\, du \\ &= \frac{t}{3}\left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}\right) \\ &\quad + \left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_x\rho\sigma_s}{\theta_1^2}\right) \\ &\quad \times \left[1 - \frac{2e^{\theta_1 t}}{\theta_1 t} - \frac{2}{(\theta_1 t)^2}(1 - e^{\theta_1 t})\right].\end{aligned} \tag{143}$$

For the term $\int_0^L \langle y_{t-L}, y_{t-L+s}\rangle\, ds$, Eq. (138) can be used for its computation. Hence, we get the last term for determining the covariance matrix of the trio (X_t, y_t, M_t) as

$$\begin{aligned}\langle M_t, M_t\rangle &= \left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}\right)\left(t - \frac{2L}{3}\right) \\ &\quad + \left[\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_x\rho\sigma_s}{\theta_1^2}\right]\left[1 - \frac{1}{(\theta_1 L)^2}\right. \\ &\quad \left. \times (1 - e^{\theta_1 L} + \theta_1 L e^{\theta_1 L}) - \frac{2}{\theta_1 L}(1 - e^{\theta_1 L})(1 - e^{\theta_1(t-L)})\right].\end{aligned} \tag{144}$$

Summarizing above, we have Lemma 1.

Lemma 1. *For $t > L$, the trio (X_t, y_t, M_t) are jointly normally distributed with mean $n = (n_1, n_2, n_3)$ given by*

$$n_1 = -\frac{\theta_0}{\theta_1}, \tag{145}$$

$$n_2 = y_0 + \left(\mu_0 - \frac{\mu_1\theta_0}{\theta_1} - \sigma_s^2/2\right)t, \tag{146}$$

$$n_3 = y_0 + \left(\mu_0 - \frac{\mu_1\theta_0}{\theta_1} - \sigma_s^2/2\right)\left(t - \frac{L}{2}\right), \tag{147}$$

and covariance matrix $D = (D_{ij})$ given by

$$D_{11} = -\frac{\sigma_x^2}{2\theta_1}, \tag{148}$$

$$\begin{aligned}D_{22} &= \left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}\right)t \\ &\quad + \left(\frac{\sigma_x^2}{\theta_1^3} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1^2}\right)(1 - e^{\theta_1 t}),\end{aligned} \tag{149}$$

$$\begin{aligned}D_{33} &= \left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_x\sigma_s\rho}{\theta_1}\right)\left(t - \frac{2L}{3}\right) \\ &\quad + \left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_s\rho\sigma_x}{\theta_1^2}\right) \\ &\quad \times \left[1 - \frac{2}{(\theta_1 L)^2}(1 - e^{\theta_1 L} + \theta_1 L e^{\theta_1 L})\right. \\ &\quad \left. - \frac{2}{\theta_1 L}(1 - e^{\theta_1 L})(1 - e^{\theta_1(t-L)})\right],\end{aligned} \tag{150}$$

$$D_{12} = \left(\frac{\mu_1\sigma_x^2}{2\theta_1^2} - \frac{\sigma_x\sigma_s\rho}{\theta_1}\right)(1 - e^{\theta_1 t}), \tag{151}$$

$$\begin{aligned}D_{13} &= \frac{1}{L}\left(-\frac{\mu_1\sigma_x^2}{2\theta_1^3} + \frac{\sigma_x\sigma_s\rho}{\theta_1^2}\right) \\ &\quad \times (1 - e^{\theta_1 L}) - \left(\frac{\mu_1\sigma_x^2}{2\theta_1^2} - \frac{\sigma_x\sigma_s\rho}{\theta_1}\right)e^{\theta_1 t},\end{aligned} \tag{152}$$

$$\begin{aligned}D_{23} &= \left(\sigma_s^2 + \frac{(\mu_1\sigma_x)^2}{\theta_1^2} - \frac{2\mu_1\sigma_s\sigma_x\rho}{\theta_1}\right)\left(t - \frac{L}{2}\right) \\ &\quad + \left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3} - \frac{\mu_1\sigma_s\rho\sigma_x}{\theta_1^2}\right)(1 - e^{\theta_1 t})\end{aligned}$$

$$-\left(\frac{(\mu_1\sigma_x)^2}{2\theta_1^3}-\frac{\mu_1\sigma_s\rho\sigma_x}{\theta_1^2}\right)\frac{1}{\theta_1 L}$$
$$\times(1-e^{\theta_1 L}-e^{\theta_1(t-L)}+e^{\theta_1 t}). \tag{153}$$

With Lemma 1, the proof of Eqs. (27) and (28) follows from Lemma 2.

Lemma 2. *Let $\hat{X}_t = X_t - \bar{X}$ and $Z_t = y_t - M_t$. Then $(\hat{X}_t, Z_t)$ is normally distributed with mean $m^Z = (n_1, n_2 - n_3)$, and covariance $C^Z = (C^Z_{ij})$ given by*

$$C^Z_{11} = D_{11}, \tag{154}$$

$$C^Z_{22} = D_{22} + D_{33} - 2D_{23}, \tag{155}$$

and

$$C^Z_{12} = D_{12} - D_{13}. \tag{156}$$

Moreover,

$$\mathrm{E}[1_{Z_t\geqslant 0}] = \mathrm{N}\left(\frac{m^Z_2}{\sqrt{C^Z_{22}}}\right) \tag{157}$$

and

$$\mathrm{E}[X_t 1_{Z_t\geqslant 0}] = m^Z_1 \mathrm{N}\left(\frac{m^Z_2}{\sqrt{C^Z_{22}}}\right) + \frac{C^Z_{12}}{\sqrt{C^Z_{22}}}\mathrm{N}'\left(-\frac{m^Z_2}{\sqrt{C^Z_{22}}}\right). \tag{158}$$

Proof. It is sufficient to prove only Eq. (158), which is generally true for any jointly normal random variable (x, z), with mean (m_x, m_z), standard deviation (σ_x, σ_z), and correlation ρ, i.e.,

$$\mathrm{E}[x1_{z\leqslant 0}] = m_x \mathrm{N}\left(\frac{m_z}{\sigma_z}\right) + \rho\sigma_x \mathrm{N}'\left(-\frac{m_z}{\sigma_z}\right). \tag{159}$$

After standardization,

$$\hat{x} = \frac{x - m_x}{\sigma_x} \tag{160}$$

and

$$\hat{z} = \frac{z - m_z}{\sigma_z}, \tag{161}$$

we can write

$$\hat{x} = \rho\hat{z} + \sqrt{1-\rho^2}\hat{e}, \tag{162}$$

where $\hat{e}$ is the standard normal variable that is independent of $\hat{z}$. Generally, for $m_z \geqslant 0$, which is satisfied by our application, where $\mathrm{E}[Z_t] = \mathrm{E}[y_t] - \mathrm{E}[M_t] > 0$. Therefore, we have

$$\begin{aligned}\mathrm{E}[x1_{z\leqslant 0}] &= \mathrm{E}[(\sigma_x\hat{x} + m_x)1_{\hat{z}\leqslant -(m_z/\sigma_z)}] \\ &= m_x \mathrm{E}1_{\hat{z}\leqslant -(m_z/\sigma_z)} + \rho\sigma_x \mathrm{E}[\hat{z}1_{\hat{z}\leqslant -(m_z/\sigma_z)}] \\ &= m_x \mathrm{N}\left(-\frac{m_z}{\sigma_z}\right) - \rho\sigma_x \mathrm{N}'\left(-\frac{m_z}{\sigma_z}\right).\end{aligned} \tag{163}$$

This implies that

$$\mathrm{E}[x1_{z\geqslant 0}] = \mathrm{E}[x] - \mathrm{E}[x1_{z\leqslant 0}] = m_x \mathrm{N}\left(\frac{m_z}{\sigma_z}\right) + \rho\sigma_x \mathrm{N}'\left(-\frac{m_z}{\sigma_z}\right), \tag{164}$$

which proves Eq. (159). □

A.2. Proof of Propositions 1–3

All three GMA strategies involve MA, which is only well defined for $t > L$. When $t \leqslant L$, we define them here as the optimal fixed strategy $\xi^*_{\text{fix}2}$, which is the same as $\xi^*_{\text{fix}1}$ under the log-utility. Thus, the complete GMA rule is

$$\mathrm{GMA}(S_t, G_t, \gamma = 1) = \begin{cases} \xi_{\text{fix}} + \xi_{\text{mv}} \cdot \eta(S_t, G_t) & \text{for } t > L, \\ \xi^*_{\text{fix}1} & \text{for } t \leqslant L. \end{cases} \tag{165}$$

This makes comparison across the strategies fair because they all start from $\xi_{\text{fix}1}$. For example, if the pure MA had started from zero, it would surely under-perform the other two over $[0, L]$ assuming a positive risk premium. Analytically, the same starting point makes the expressions simpler. Clearly, for a fixed L, the initial value has little impact, if any, when T is large. This is also consistent with the numerical results in Section 4.1. However, when study optimal lags, the initial value does matter because the optimal lag of a pure MA strategy can be close to T (see Section 4.4).

With any of the GMA strategies, the key is to maximize the expected log-utility, which follows from Appendix A.1 and Eq. (32), as a function of ξ_{fix} and ξ_{mv},

$$\begin{aligned} U_{\text{GMA}}(\xi_{\text{fix}}, \xi_{\text{mv}}) = {} & \log W_0 + rT + \frac{(\mu_0 + \mu_1\bar{X} - r)^2}{2\sigma_s^2}L \\ & + \xi_{\text{fix}}\left[\mu_0 + \mu_1\bar{X} - r - \frac{\sigma_s^2}{2}\xi_{\text{fix}}\right](T - L) \\ & + \xi_{\text{mv}}\mu_1 b_1(T-L) + \left[\xi_{\text{mv}}(\mu_0 + \mu_1\bar{X} - r)\right. \\ & \left. - \frac{\sigma_s^2}{2}\xi^2_{\text{mv}} - \sigma_s^2\xi_{\text{fix}}\xi_{\text{mv}}\right]b_2(T-L), \end{aligned} \tag{166}$$

where b_1 and b_2 are defined in Eqs. (27) and (28).

To prove Proposition 1, we need to maximize $U_{\text{GMA}}(\xi_{\text{fix}}, \xi_{\text{mv}})$ with respect to both ξ_{fix} and ξ_{mv}. The first-order conditions are

$$\left.\frac{\partial U_{\text{GMA}}(\xi_{\text{fix}}, \xi_{\text{mv}})}{\partial \xi_{\text{fix}}}\right|_{\xi_{\text{fix}}=\xi^*_{\text{fix}}, \xi_{\text{mv}}=\xi^*_{\text{mv}}} = 0 \tag{167}$$

and

$$\left.\frac{\partial U_{\text{GMA}}(\xi_{\text{fix}}, \xi_{\text{mv}})}{\partial \xi_{\text{mv}}}\right|_{\xi_{\text{fix}}=\xi^*_{\text{fix}}, \xi_{\text{mv}}=\xi^*_{\text{mv}}} = 0, \tag{168}$$

which implies

$$\mu_0 + \mu_1\bar{X} - r - \sigma_s^2\xi_{\text{fix}} - \sigma_s^2\xi_{\text{mv}}b_2 = 0 \tag{169}$$

and

$$b_1 + (\mu_0 + \mu_1\bar{X} - r)b_2 - \sigma_s^2(\xi_{\text{fix}} + \xi_{\text{mv}})b_2 = 0. \tag{170}$$

With some algebra, we obtain the optimal solution

$$\xi^*_{\text{fix}} = \frac{\mu_0 + \mu_1\bar{X} - r}{\sigma_s^2} - \frac{\mu_1 b_1}{(1 - b_2)\sigma_s^2} \tag{171}$$

and

$$\xi^*_{\text{mv}} = \frac{\mu_1 b_1}{b_2(1 - b_2)\sigma_s^2}. \tag{172}$$

Because the value function for log-utility associated with ξ^*_{fix1} is

$$U^*_{\text{fix1}} = \log W_0 + rT + \frac{(\mu_0 + \mu_1\bar{X} - r)^2}{2\sigma_s^2}T, \tag{173}$$

we obtain Eq. (35) by substituting this into $U_{\text{GMA}}(\xi_{\text{fix}}, \xi_{\text{mv}})$ evaluated at the optimal solution $(\xi^*_{\text{fix}}, \xi^*_{\text{mv}})$.

To prove Proposition 2, we simply let $\xi_{\text{fix}} = \xi^*_{\text{fix1}}$ and optimize $U_{\text{GMA}}(\xi^*_{\text{fix1}}, \xi_{\text{mv}})$ over ξ_{mv} alone. Similar algebra yields the solution. The proof of Proposition 3 follows analogously.

A.3. Proof of Eq. (64)

To maximize $U(\gamma)$ of Eq. (59) over ξ_{mv}, it is equivalent to maximize

$$\max_{\xi_{\text{mv}}} f(\xi_{\text{mv}}) = \xi_{\text{mv}}(\phi_0 + \phi_1\xi_{\text{mv}} + \phi_2\xi^2_{\text{mv}} + \phi_3\xi^3_{\text{mv}}). \tag{174}$$

The first-order condition is

$$f'(\xi_{\text{mv}}) = \phi_0 + 2\phi_1\xi_{\text{mv}} + 3\phi_2\xi^2_{\text{mv}} + 4\phi_3\xi^3_{\text{mv}} = 0, \tag{175}$$

which in turn can be transformed to

$$y^3 + py + q = 0, \tag{176}$$

where

$$y = \xi_{\text{mv}} + \frac{\phi_2}{4\phi_3} \tag{177}$$

with p and q given in Eqs. (65) and (66). Numerical computations show that, for a wide range of parameters of interest, we have

$$q^2 + \frac{4p^3}{27} > 0. \tag{178}$$

The solution to cubic equation (176) is known as Cardano solution (e.g., Curtis, 1944), which is given by

$$y^* = -\left[\frac{q + \sqrt{q^2 + 4p^3/27}}{2}\right]^{1/3} + \frac{p}{3}\left[\frac{q + \sqrt{q^2 + 4p^3/27}}{2}\right]^{-1/3}. \tag{179}$$

Under condition equation (178), this is the unique real root. Hence

$$\xi^*_{\text{mv}} = -\frac{\phi_2}{4\phi_3} + y^*, \tag{180}$$

which is the same as Eq. (64). Furthermore, it can be verified that $\phi_1 < 0$, and so this solution to Eq. (175) is a maximum.

A.4. Proof of Proposition 4

To prove Proposition 4, we need to optimize Eqs. (35), (42) and (45) over L. Consider $U^*_{\text{GMA1}} - U^*_{\text{fix1}}$ and $U^*_{\text{GMA2}} - U^*_{\text{fix1}}$, and ignore some constants, the target functions become

$$U_1 = \frac{b_1^2}{b_2(1-b_2)}\left(1 - \frac{L}{T}\right) = V_1\left(1 - \frac{L}{T}\right) \tag{181}$$

and

$$U_2 = \frac{b_1^2}{b_2}\left(1 - \frac{L}{T}\right) = V_2\left(1 - \frac{L}{T}\right), \tag{182}$$

where V_1 and V_2 are defined accordingly. Because V_1 and V_2 are T independent, so are their maximum over L. As T is large, $1 - L/T$ can be ignored, and hence we need only to maximize V_1 and V_2.

The first-order condition for maximizing V_2 is

$$V_2' = \frac{2b_1b_1'b_2 - b_1^2b_2'}{b_2^2} = 0. \tag{183}$$

Substituting those approximate equations (105) and (106) for b_1 and b_2, we have

$$2h'(x)f(Ax) - 2Axh(x)f(Ax) - \frac{Ah(x)f^2(Ax)}{\text{N}(Ax)} = 0. \tag{184}$$

This is a transcendant equation that is difficult to solve without further simplifications. It can be shown that the third term is dominated by the first one when $x < 1$, and by the second one when $x > 1$. Ignoring the third term, we need only to optimize

$$b_1 = h(x) \cdot f(Ax). \tag{185}$$

The Taylor expansion for $h(x)$ is

$$h(x) = \frac{x}{2} - \frac{x^3}{6} + \frac{x^5}{24} + \cdots, \tag{186}$$

which implies that Eq. (185) can be approximated by

$$\left(\frac{x}{2} - \frac{x^3}{6} + \frac{x^5}{24}\right)\exp\left(-\frac{A^2x^2}{2}\right). \tag{187}$$

Taking derivative with respect to x and letting it be equal to zero, we obtain, after ignoring higher-order terms,

$$\left(\frac{5}{24} + \frac{A^2}{6}\right)x^4 - \frac{1 + A^2}{2}x^2 + \frac{1}{2} = 0. \tag{188}$$

The smaller root of the above quadratic equation, which corresponds to the maximum, is the solution for the second case of Proposition 4.

To provide solution for the first case, we now maximize V_1. Its denominator can be approximated by $\text{N}(Ax)\cdot\text{N}(-Ax)$, and hence

$$V_1 \approx \frac{h^2(x)f^2(Ax)}{\text{N}(Ax)\text{N}(-Ax)} = \frac{1}{C\cdot\text{N}(Ax)}[h(x)\sqrt{f(Ax)}]^2, \tag{189}$$

where we have used the approximation $\text{N}(-Ax) \approx C \cdot f(Ax)$ for $Ax > 0$ and large. Similar to the earlier case, we can ignore $\text{N}(Ax)$, and hence the target function becomes $h(x)\cdot\sqrt{f(Ax)}$. This has the same form as Eq. (185) with $A/\sqrt{2}$ plays the role of earlier A. Therefore, the solution follows.

Finally, to derive Eq. (111), we need to maximize $U_3 = U^*_{GMA3} - ((\mu_s - r)^2/2\sigma_s^2)L$. Similarly, this can be replaced by a target function

$$V_3 = [\mu_1 C_4 h(x) f(Ax) + C_5 N(Ax)] \cdot \left(1 - \frac{x^2}{|\theta_1|T}\right)$$
$$= \left[\mu_1 C_4 \cdot \frac{1}{x}\left(1 - \frac{1-e^{-x^2}}{x^2}\right) f(Ax) + C_5 N(Ax)\right] \cdot \left(1 - \frac{x^2}{|\theta_1|T}\right)$$
$$\approx C_5 N(Ax) \cdot \left(1 - \frac{x^2}{|\theta_1|T}\right), \tag{190}$$

where the last approximation is due to the dominance of the second term in the bracket. The first-order condition is

$$f(Ax) \cdot \left(1 - \frac{x^2}{|\theta_1|T}\right) - \frac{2}{|\theta_1|T} x N(Ax) = 0. \tag{191}$$

Because there is only one solution, we can verify that

$$|\theta_1|T \gg 1, \tag{192}$$

$$Ax \gg 1, \tag{193}$$

and

$$\frac{x^2}{|\theta_1|T} \to 0, \tag{194}$$

and hence we can reduce the first-order condition to $Af(Ax) \approx (2/|\theta_1|T)x$. This implies Eq. (111).

A.5. Computing the ML estimators

Following Huang and Liu (2007), the continuously compounded return $R_{t+1} = \log(S_{t+1}/S_t)$ and X_{t+1} are jointly Gaussian, and the log-likelihood function, conditional on X_0, can be written as

$$\pounds(\Theta) = \sum_{t=1}^{T} \log f(R_t, X_t | X_{t-1}; \Theta)$$
$$= -\frac{T}{2}(2\log 2\pi + \log\sigma_1^2 + \log\sigma_2^2 + \log(1-\rho_{12}^2))$$
$$- \frac{1}{2(1-\rho_{12}^2)} \sum_{t=1}^{T} \left\{ \frac{(R_t - a_{11} - a_{12}X_{t-1})^2}{\sigma_1^2} + \frac{(X_t - b_{11} - b_{12}X_{t-1})^2}{\sigma_2^2} - \frac{2\rho_{12}(R_t - a_{11} - a_{12}X_{t-1})(X_t - b_{11} - b_{12}X_{t-1})}{\sigma_1\sigma_2} \right\}, \tag{195}$$

where $\Theta \equiv (a_{11}, a_{12}, b_{11}, b_{12}, \sigma_1, \sigma_2, \rho_{12})$ with

$$a_{11} = \left(\mu_0 - \frac{1}{2}\sigma_s^2 - \frac{\mu_1\theta_0}{\theta_1}\right)\Delta t + \frac{\mu_1\theta_0}{\theta_1^2}(e^{\theta_1\Delta t} - 1), \tag{196}$$

$$a_{12} = \frac{\mu_1}{\theta_1}(e^{\theta_1\Delta t} - 1), \tag{197}$$

$$b_{11} = \frac{\theta_0}{\theta_1}(e^{\theta_1\Delta t} - 1), \tag{198}$$

$$b_{12} = e^{\theta_1\Delta t}, \tag{199}$$

$$\sigma_1^2 = \left(\sigma_s^2 + \frac{\mu_1^2}{\theta_1^2}\sigma_x^2 - \frac{2\mu_1}{\theta_1}\rho\sigma_s\sigma_x\right)\Delta t + \frac{1}{2\theta_1}(e^{2\theta_1\Delta t} - 1)\frac{\mu_1^2}{\theta_1^2}\sigma_x^2 + \frac{2\mu_1}{\theta_1^2}(e^{\theta_1\Delta t} - 1) \times \left(\rho\sigma_s\sigma_x - \frac{\mu_1}{\theta_1}\sigma_x^2\right), \tag{200}$$

$$\sigma_2^2 = \frac{\sigma_x^2}{2\theta_1}(e^{2\theta_1\Delta t} - 1), \tag{201}$$

and

$$\rho_{12}\sigma_1\sigma_2 = \frac{\mu_1}{2\theta_1^2}(e^{\theta_1\Delta t} - 1)^2\sigma_x^2 + \frac{\rho\sigma_s\sigma_x}{\theta_1}(e^{\theta_1\Delta_t} - 1). \tag{202}$$

Let Y be a $T \times 2$ matrix formed by observation on R_t and X_t, and Z be formed by a T-vector of ones and the T values of X_{t-1}. Define

$$B = \begin{pmatrix} a_{11} & b_{11} \\ a_{12} & b_{12} \end{pmatrix} \tag{203}$$

and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{204}$$

Then, the estimator of B is $\hat{B} = (X'X)^{-1}X'Y$, and that of Σ is $\hat{\Sigma} = (Y - X\hat{B})'(Y - X\hat{B})/T$. The estimator for the original parameters, such as μ_0, can be backed out from these estimates.

A.6. The linear rule

Aït-Sahalia and Brandt (2001) and especially Brandt and Santa-Clara (2006) provide linear portfolio rules to approximate the optimal dynamic strategy. Following Brandt and Santa-Clara (2006), consider linear portfolio rules of the following form:

$$\xi_t = \xi_{0,t+j} + \xi_{1,t+j}X_{t+j}, \quad j = 1, \ldots, H, \tag{205}$$

where H is the investment horizon. Their idea is to reduce the multi-period problem to a single-period one by expanding the asset space with "conditional managed portfolio" and "timing portfolio" according to Eqs. (12) and (25) in their paper. Our model has one risk-free asset and one risky asset. Denote here $R_f = 1 + r_f\Delta t$ as the gross return on the risk-free asset and $r_t = R_t - R_f$ the excess returns on the risky asset. Then, the expanded asset space can be written as

$$\tilde{r}'_{t\to t+H} = [\{R_f^{H-1} r_{t+j+1}\}_{j=0}^{H-1}, \{R_f^{H-1} X_{t+j} r_{t+j+1}\}_{j=0}^{H-1}], \tag{206}$$

which is an $1 \times 2H$ vector.

The multi-period utility maximization problem can thus be approximated by

$$\max_{\theta_t} E_t[u(R_f^H + \theta'_t \tilde{r}_{t\to t+H})], \tag{207}$$

where θ'_t, $1 \times 2H$, is the single-period portfolio position in the expanded asset space. To solve this problem, Brandt and Santa-Clara (2006) suggest a further approximation by replacing the power utility with its fourth-order expansion, i.e.,

$$\begin{aligned}E_t[u(W_{t+H})] \approx E_t[&u(W_tR_f^H) + u'(W_tR_f^H)(W_t\theta_t'\tilde{r}_{t\to t+H})\\&+\tfrac{1}{2}u''(W_tR_f^H)(W_t\theta_t'\tilde{r}_{t\to t+H})^2\\&+\tfrac{1}{6}u'''(W_tR_f^H)(W_t\theta_t'\tilde{r}_{t\to t+H})^3\\&+\tfrac{1}{24}u''''(W_tR_f^H)(W_t\theta_t'\tilde{r}_{t\to t+H})^4].\end{aligned} \tag{208}$$

As a result,

$$\begin{aligned}\theta_t' \approx &-\{E_t[u''(W_tR_f^H)(\tilde{r}_{t\to t+H}\tilde{r}_{t\to t+H}')]\}^{-1}\\&\times\{E_t[u'(W_tR_f^H)(\tilde{r}_{t\to t+H})]W_t\\&+\tfrac{1}{2}E_t[u'''(W_tR_f^H)(\theta_t'\tilde{r}_{t\to t+H})^2\tilde{r}_{t\to t+H}]W_t^3\\&+\tfrac{1}{6}E_t[u''''(W_tR_f^H)(\theta_t'\tilde{r}_{t\to t+H})^3\tilde{r}_{t\to t+H}]W_t^4\}.\end{aligned} \tag{209}$$

Based on the predictive model, the above moments can be computed via simulations, and hence the implicit expression for the optimal weights can be solved recursively.

References

Aït-Sahalia, Y., Brandt, M., 2001. Variable selection for portfolio choice. Journal of Finance 56, 1297–1351.
Allen, F., Karjalainen, R., 1999. Using genetic algorithms to find technical trading rules. Journal of Financial Economics 51, 245–271.
Ang, A., Bekaert, G., 2006. Stock return predictability: Is it there? Review of Financial Studies 20, 651–707.
Back, K., 2006. Introduction to asset pricing and portfolio choice theory. Unpublished manuscript, Texas A&M University, Texas.
Barberis, N., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.
Bhattacharya, U., Galpin, N., 2006. Is stock picking declining around the world? Unpublished working paper, Indiana University, Bloomington, Indiana.
Blume, L., Easley, D., O'Hara, M., 1994. Market statistics and technical analysis: the role of volume. Journal of Finance 49, 153–181.
Brandt, M., Santa-Clara, P., 2006. Dynamic portfolio selection by augmenting the asset space. Journal of Finance 61, 2187–2217.
Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance 47, 1731–1764.
Brown, D., Jennings, R., 1989. On technical analysis. Review of Financial Studies 2, 527–551.
Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.
Campbell, J.Y., Thompson, S.B., 2008. Predicting the equity premium out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.
Campbell, J.Y., Viceira, L.M., 1999. Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics 114, 433–495.
Cochrane, J.H., 2008. The dog that did not bark: a defense of return predictability. Review of Financial Studies 21, 1533–1575.
Covel, M., 2005. Trend Following: How Great Traders Make Millions in Up or Down Markets. Prentice-Hall, New York.
Cowles, A., 1933. Can stock market forecasters forecast? Econometrica 1, 309–324.
Curtis, H., 1944. A derivation of Cardano's formula. American Mathematical Monthly 51, 35.
Fama, E.F., Blume, M., 1966. Filter rules and stock market trading. Journal of Business 39, 226–241.
Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. Journal of Finance 25, 383–417.
Fama, E., Schwert, G., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.
Ferson, W., Harvey, C., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385–415.
Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.
Gehrig, T., Menkhoff, L., 2006. Extended evidence on the use of technical analysis in foreign exchange. International Journal of Finance and Economics 11, 327–338.
Gennotte, G., 1986. Optimal portfolio choice under incomplete information. Journal of Finance 41, 733–746.
Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Management Science 49, 639–654.
Grundy, B., Kim, Y., 2002. Stock market volatility in a heterogeneous information economy. Journal of Financial and Quantitative Analysis 37, 1–27.
Huang, L., Liu, H., 2007. Rational inattention and portfolio selection. Journal of Finance 62, 1999–2040.
Ingersoll, J., 1987. Theory of Financial Decision Making. Rowman and Littlefield, New York.
Jensen, M.C., Benington, G.A., 1970. Random walks and technical theories: some additional evidence. Journal of Finance 25, 469–482.
Kandel, S., Stambaugh, R., 1996. On the predictability of stock returns: an asset-allocation perspective. Journal of Finance 51, 385–424.
Karatzas, I., Shreve, S., 1991. Brownian Motion and Stochastic Calculus, second ed. Springer, New York.
Kim, T., Omberg, E., 1996. Dynamic nonmyopic portfolio behavior. Review of Financial Studies 9, 141–161.
Kirkpatrick, C.D., Dahlquist, J.R., 2006. Technical Analysis: The Complete Resource for Financial Market Technicians. Prentice-Hall, New York.
LeBaron, B., 1999. Technical trading rule profitability and foreign exchange intervention. Journal of International Economics 49, 125–143.
Litterman, B., 2003. Modern Investment Management. Wiley, New York.
Liu, J., 2007. Portfolio selection in stochastic environments. Review of Financial Studies 20, 1–39.
Lo, A., MacKinlay, C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies 1, 41–66.
Lo, A., MacKinlay, C., 1990. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3, 431–468.
Lo, A., MacKinlay, C., 1999. A Non-Random Walk Down Wall Street. Princeton University Press, Princeton, NJ.
Lo, A., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. Journal of Finance 55, 1705–1765.
Magdon-Ismail, M., Atiya, A., Pratap, A., Abu-Mostafa, Y., 2004. On the maximum drawdown of a Brownian motion. Journal of Applied Probability 41, 147–161.
Malkiel, B., 1981. A Random Walk Down Wall Street. Norton, New York.
Markowitz, H., 1952. Mean–variance analysis in portfolio choice and capital markets. Journal of Finance 7, 77–91.
Merton, R., 1992. Continuous-time Finance. Blackwell, Cambridge, MA.
Meucci, A., 2005. Risk and Asset Allocation. Springer, New York.
Murphy, J., 1986. Technical Analysis of Futures Markets. New York Institute of Finance, New York.
Neely, C.J., 2002. The temporal pattern of trading rule returns and exchange rate intervention: intervention does not generate technical trading profits. Journal of International Economics 58, 211–232.
Nison, S., 1991. Japanese Candlestick Charting Techniques. Simon and Schuster, New York.
Pástor, Ľ., Stambaugh, R., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335–381.
Rapach, D.E., Strauss, J., Zhou, G., 2009. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Unpublished manuscript, Washington University, St. Louis, Missouri.
Rubinstein, M., 2002. Markowitz's "portfolio selection": a 50-year retrospective. Journal of Finance 57, 1041–1045.
Schwager, J., 1993. Market Wizards: Interviews with Top Traders. Collins, New York.
Schwager, J., 1995. The New Market Wizards: Conversations with America's Top Traders. Wiley, New York.
Schwert, B., 2003. Anomalies and market efficiency. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance. North-Holland, Amsterdam, pp. 937–972.
Shefrin, H., 2008. A Behavioral Approach to Asset Pricing. Academic Press, New York.
Shleifer, A., 2000. Inefficient Markets: An Introduction to Behavioral Finance. Oxford University Press, New York.
Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.
Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 53, 1647–1691.
Tobin, J., 1958. Liquidity preference as behavior towards risk. Review of Economic Studies 25, 65–86.

Tu, J., Zhou, G., 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? Journal of Financial Economics 72, 385–421.

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

Xia, Y., 2001. Learning about predictability: the effects of parameter uncertainty on dynamic asset allocation. Journal of Finance 55, 205–246.

Zhou, G., Zhu, Y., 2009. Asset allocation: Is technical analysis useful? Unpublished manuscript, Washington University, St. Louis, Missouri.