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

Yingzi Zhu, Guofu Zhou

School of Economics and Management, Tsinghua University, Beijing 100084, China
Olin Business School, Washington University in St. Louis, St. Louis, MO 63130, USA
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:
G12
G14
C11
C61

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.

© 2009 Elsevier B.V. All rights reserved.

Keywords:
Technical analysis
Trading rules
Asset allocation
Predictability
Learning

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...
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. 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 \( L \). 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 approximately 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.

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, Gennette (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 underperforms 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. Behavioral 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., Cowles, 2005; Kirkpatrick and Dahquist, 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. 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 \( S_t \):

\[
\frac{dS_t}{S_t} = (\mu_0 + \mu_1X_t) dt + \sigma_1 dB_t
\]

and

\[
dX_t = (\theta_0 + \theta_1X_t) dt + \sigma_2 dZ_t,
\]

where \( \mu_0, \mu_1, \sigma_1, \theta_0, \theta_1, \) and \( \sigma_2 \) are parameters; \( X_t \) is a predictive variable; and \( B_t \) and \( Z_t \) are standard Brownian motions with correlation coefficient \( \rho \); \( \theta_1 \) 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 \( \zeta_t \) to maximize his expected utility of wealth.

\[
\max_{\zeta_t} \mathbb{E}[u(W_T)],
\]

subject to the budget constraint

\[
\frac{dW_t}{W_t} = r dt + \zeta_t(\mu_0 + \mu_1X_t - r) dt + \zeta_t\sigma_1 dB_t.
\]

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

\[
u(W_t) = \frac{W_t^{-\gamma} - 1}{1 - \gamma},
\]

where \( \gamma \) 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

\[
\zeta_t^* = \frac{\mu_0 + \mu_1X_t - r}{\gamma \sigma_1^2} + \frac{(1 - \gamma)\rho \sigma_1}{\gamma \sigma_2} [\chi(t) + \zeta(t)X_t],
\]

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

\[
\chi(t) + a_1\chi(t) + a_2\chi(t) + a_4\zeta(t) + a_5 = 0
\]

and

\[
\zeta(t) + a_1\zeta(t) + a_2\zeta(t) + a_3 = 0,
\]

with

\[
a_1 = \frac{(1 - \gamma)^2}{\gamma} \rho^2 \sigma_2 + (1 - \gamma)\sigma_2^2,
\]

\[
a_2 = 2\left(1 - \gamma \frac{\mu_1}{\gamma} \rho^2 \sigma_2^2 + \theta_1\right),
\]

\[
a_3 = \frac{1}{\gamma} \left(\frac{\mu_1}{\gamma}\right)^2,
\]

\[
a_4 = \frac{1 - \gamma \theta_0 - r}{\gamma \sigma_2},
\]

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
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

\[ \zeta_{\text{fix}} = \frac{\mu - r}{\sigma^2}, \]

where \( \mu \) is the long-term mean of the stock return. This strategy invests a fixed or constant proportion of wealth, \( \zeta_{\text{fix}} \), 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 \( \zeta_{\text{fix}} \) 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 \( \zeta \) to a constant. The solution is analytically obtained as

\[ \zeta_{\text{fix}2} = \frac{\mu - r}{\gamma \sigma^2 + (1 - \gamma)(A + 2\mu_1 \sigma_1 b)}, \]

where

\[ A = \frac{\sigma_1^2}{\theta_1} \left( 1 + \frac{1 - e^{\theta_1 T}}{\theta_1 T} \right), \]

and

\[ B = \frac{\rho \sigma_1}{\theta_1} \left( \frac{e^{\theta_1 T} - 1}{\theta_1 T} - 1 \right). \]

For \( \gamma = 1 \), this optimal constant strategy is equal to \( \zeta_{\text{fix}} \). 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

\[ A_t = \frac{1}{L} \int_{t-L}^t S_u \, du, \]

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

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

This is well defined when \( t > L \) and can be taken as zero or as another fixed constant when \( t < L \). 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 \( \eta_t \). This is clearly not optimal for two reasons. First, the MA rule should in general be a function of the risk-aversion parameter \( \gamma \). Intuitively, \( \gamma \) 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,

\[ \text{GMA}(S_t, A_t, \gamma) = \zeta_{\text{fix}} + \xi_{\text{mv}} \cdot \eta(S_t, A_t), \]

where \( \zeta_{\text{fix}} \) and \( \xi_{\text{mv}} \) 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, \( \zeta_{\text{fix}} \) is obtained by setting \( \zeta_{\text{fix}} = 0 \) and \( \xi_{\text{mv}} = 0 \), and \( \eta_t \) is obtained by setting \( \zeta_{\text{fix}} = 1 \) and \( \xi_{\text{mv}} = 1 \).

Three interesting questions are associated with the GMA rule. First, what is the optimal choice of \( \zeta_{\text{fix}} \) and \( \xi_{\text{mv}} \) and how well does it perform compared with other fixed strategies? Second, with \( \zeta_{\text{fix}} \) being equal to either \( \zeta_{\text{fix}} \) or \( \zeta_{\text{fix}2} \), is the optimal choice of \( \xi_{\text{mv}} \) 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 \( \zeta_{\text{fix}} = 0 \), what is the optimal choice of \( \xi_{\text{mv}} \)? 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 \( \xi_{\text{mv}} \) 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

---


\(^7\) 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.
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
\[
dW_t = [r + \text{GMA} \cdot (\mu_0 + \mu_1 X_t - r)] dt + \text{GMA} \cdot \sigma_d dB_t
\]
and, hence, assuming \( T > L \), we have
\[
\log W_T = \log W_0 + rT + \int_0^T dt \left[ \xi_{\text{fix}} \left( \mu_0 + \mu_1 X_t - r - \frac{\sigma_2^2}{2} \xi_{\text{fix}} \right) \right] + \int_0^T dt \left[ \xi_{\text{mv}} \left( \mu_0 + \mu_1 X_t - r - \frac{\sigma_2^2}{2} \xi_{\text{mv}} \right) \right] + \xi_{\text{mv}} \mu_1 \int_0^T dt \left[ \xi_{\text{fix}} \left( \mu_0 + \mu_1 X_t - r \right) \right] - \frac{\sigma^2_2}{2} \xi_{\text{mv}} - \frac{\sigma_2^2}{2} \xi_{\text{fix}} \xi_{\text{mv}} \eta_t + \sigma_{\text{mv}} \int_0^T \xi_{\text{mv}} dB_t,
\]
where \( X_t = X_t - \bar{X} \) with \( \bar{X} = -\theta_0 / \theta_1 \). Assume that \( X_t \) is stationary, and it starts from its steady state distribution. Then, the expected log-utility is
\[
U_{\text{GMA}} = E \log W_T = \log W_0 + rT + \left( \mu_0 + \mu_1 \bar{X} - r \right)^2 L + \int_0^T dt \xi_{\text{fix}} \left[ \mu_0 + \mu_1 \bar{X} - r - \frac{\sigma_2^2}{2} \xi_{\text{fix}} \right] + \int_0^T dt \xi_{\text{mv}} \mu_1 \xi_{\text{fix}} \eta_t + \int_0^T dt \xi_{\text{mv}} \left( \mu_0 + \mu_1 \bar{X} - r \right) - \frac{\sigma^2_2}{2} \xi_{\text{mv}} - \xi_{\text{fix}} \xi_{\text{mv}} \xi_{\text{mv}} \xi_{\text{fix}} \xi_{\text{mv}} \eta_t,
\]
where \( \eta_t \) is the covariance between \( X_t \) and the MA strategy \( \eta_t \) and \( b_2 \) is the probability of \( S_t > G_t \) at any given time. It can be shown that
\[
b_1 = \frac{E \xi_{\text{fix}} (S_t, G_t)}{\sqrt{C_{12}^2}} \frac{C_{12}}{\sqrt{C_{22}^2}} N \left( \frac{m_{\text{fix}}}{\sqrt{C_{12}^2}} \right)
\]
and
\[
b_2 = \frac{E \eta (S_t, G_t)}{\sqrt{C_{12}^2}} N \left( \frac{m_{\text{fix}}}{\sqrt{C_{12}^2}} \right).
\]

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 GMA(S_t, G_t, \gamma), the optimal choice of \( \xi_{\text{fix}} \) and \( \xi_{\text{mv}} \) under the log-utility is
\[
\xi_{\text{fix}}^* = \frac{m_{\text{fix}}}{\sigma_{\text{fix}}^2} \left( 1 - b_2 \sigma_{\text{fix}}^2 \right),
\]
and
\[
\xi_{\text{mv}}^* = \frac{m_{\text{fix}} b_1}{b_2 (1 - b_2 \sigma_{\text{fix}}^2)}.
\]
and the associated value function is

\[ U_{\text{GMA1}}^* = U_{\text{fix}}^* + \frac{\mu^2 b_1^2}{2 b_2 (1 - b_2) \sigma_f^2} (T - L) \geq U_{\text{fix}}^*. \tag{35} \]

where \( U_{\text{fix}}^* \) is the value function associated with \( \xi_{\text{fix}}^* \).

Proposition 1 says that the improvement over \( \xi_{\text{fix}}^* \) 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 \( \xi_{\text{fix}}^* \) is optimal already. The point is that \( \xi_{\text{fix}}^* \) 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, \( \xi_{\text{fix}}^* \) is always positive by combining a suitable fixed strategy with the MA one unless \( \sigma_f^2 \) 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 \( \xi_{\text{fix}}^* \) to add value. To address this issue, we need to solve the earlier optimization by imposing the constraint that \( \xi_{\text{fix}} = \xi_{\text{fix}}^* \). In this case, we present Proposition 2.

Proposition 2. In the class of strategies GMA \((S_1, G, \gamma)\) with \( \xi_{\text{fix}} \) being set at \( \xi_{\text{fix}}^* \), the optimal choice of \( \xi_{\text{inv}} \) under the log-utility is

\[ \xi_{\text{inv}}^* = \frac{\mu_1 b_1}{b_2 \sigma_f^2}; \tag{41} \]

and the associated value function is

\[ U_{\text{GMA2}} = U_{\text{fix}}^* + \frac{\mu^2 b_1^2}{2 b_2 \sigma_f^2} (T - L) \geq U_{\text{fix}}^*. \tag{42} \]

where \( U_{\text{fix}}^* \) is the value function associated with \( \xi_{\text{fix}}^* \).

As for \( U_{\text{GMA1}}^* \), \( U_{\text{GMA2}} \) is at least as large as \( U_{\text{fix}}^* \). When there is predictability, \( U_{\text{GMA2}} \) is clearly strictly larger than \( U_{\text{fix}}^* \), 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 \( \xi_{\text{inv}}^* \) in Proposition 2 differs from that in Proposition 1 by only a factor of \( 1 - b_2 \) in the denominator. Because \( 0 < b_2 < 1 \), \( \xi_{\text{inv}}^* \) is smaller now in absolute value. This is expected. Because \( \xi_{\text{fix}}^* \) is set at \( \xi_{\text{fix}}^* \), the risk exposure to the stock market is relatively higher already as \( \xi_{\text{inv}}^* \geq \xi_{\text{inv}}^* \). Hence, when the MA rule detects an upward trend in the market, the investor acts more aggressively than \( \xi_{\text{fix}}^* \), but less aggressively than \( \xi_{\text{fix}}^* \), before finally, it is seen that

\[ U_{\text{GMA2}} = U_{\text{GMA1}} - \frac{\mu^2 b_1^2}{2 (1 - b_2) \sigma_f^2} (T - L) \leq U_{\text{GMA1}} \leq U_{\text{opt}}. \tag{43} \]

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 \( \xi_{\text{fix}}^* \).

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 GMA \((S_1, G, \gamma)\) with restriction \( \xi_{\text{fix}} = 0 \), the optimal choice of \( \xi_{\text{inv}} \) under
the log-utility is

$$\xi^*_{\text{mv}} = -\frac{\mu_t - r}{\sigma_t^2} + \frac{\mu_t b_1}{\sigma_t^2 b_2^2},$$  \hspace{1cm} (44)$$

and the associated value function is

$$U^*_{\text{GMA3}} = U^*_{\text{fix1}} + \left(\mu_1 b_1 + (\mu_2 - r)b_2^2 - (\mu_4 - r)b_2^2\right) (T - L),$$  \hspace{1cm} (45)$$

which can be either greater or smaller than \(U^*_{\text{fix1}}\), the value function associated with \(\xi^*_{\text{fix1}}\).

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 allocating capital, say \(\xi_1\), based on a standard fixed allocation model and then, in the second step, apply this in the market-timing decision: invest more than \(\xi_1\) 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

$$\mu_t - r < -\frac{\mu_1 b_1}{\sqrt{b_2} - b_2},$$  \hspace{1cm} (46)$$

the relation about the risk premium, is satisfied, then the pure MA strategy does yield a higher expected utility than the fixed strategy \(\xi^*_{\text{fix1}}\). 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 \(\gamma\), the investor's risk aversion.

By approximating \(\int X_t \, dt\), \(\int X_t \mu_t \, dt\) and \(\int \mu_t \, dt\) with their mean values, we can write the expected utility under the GMA as

$$U_{\text{GMA}}(\gamma) \approx \frac{\mathcal{W}_0 \exp(r T) \gamma^y}{1 - \gamma} \left[ \left(1 - y\right) \left[ \xi_{\text{fix1}}(u_0 + \mu_t X_t - r) - \frac{\gamma^2}{2} z^2 \right] \right]$$

$$+ \frac{\xi_{\text{mv2}}}{2} \mathcal{E}(X_t, \eta_t) + \left[ \xi_{\text{mv1}}(u_0 + \mu_t X_t - r) - \frac{\gamma^2}{2} z^2 \right] \left[ \mathcal{E}(X_t, \eta_t) \right].$$  \hspace{1cm} (47)$$

Optimizing this approximated utility function, we obtain

$$U_{\text{GMA}}(\gamma) = \frac{1}{\gamma} GMA(S_t, G_t, 1).$$  \hspace{1cm} (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

$$\xi^*_{\text{fix1}} = \frac{\mu_t - r}{\gamma^2 z^2} - \frac{\mu_1 b_1}{\gamma (1 - b_2) \sigma_t^2},$$  \hspace{1cm} (49)$$

and

$$\xi^*_{\text{mv1}} = \frac{\mu_1 b_1}{\gamma b_2 (1 - b_2) \sigma_t^2}.$$  \hspace{1cm} (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 \(\gamma\), 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,

$$U(\gamma) = \frac{1}{1-\gamma}E[\psi(\gamma)] = \frac{1}{1-\gamma}E[\exp((1-\gamma)\log W_t)]$$

(51)
can be approximated by

$$U(\gamma) = \frac{(W_0 \exp(\gamma T))^{1-\gamma}}{1-\gamma}U_{\text{fin}}(\xi_{\text{fin}})$$

$$\exp\left\{\left(1-\gamma\right)\xi_{\text{fin}} E[C_T + D_T + \gamma(\xi_{\text{fin}}, \xi_{\text{inv}}) F_T]\right\}$$

$$+ \frac{1}{2}\left(1-\gamma\right)^2 \xi_{\text{inv}} ^2 \text{var}(C_T + D_T + \gamma(\xi_{\text{fin}}, \xi_{\text{inv}}) F_T)$$

$$+ \left(1-\gamma\right)\xi_{\text{fin}} \xi_{\text{inv}} \text{cov}(A_T + B_T, C_T + D_T + y F_T)\right\}.$$ 

(52)

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

$$\psi(\gamma) = (\mu_0 + \mu_1 X - \frac{1}{2}\sigma_1^2 \xi_{\text{inv}} - \sigma_2^2 \xi_{\text{fin}})$$

(53)

and

$$\xi_{\text{fin}} = \frac{\phi_0}{3\phi_3 - 1} - \frac{2\phi_2}{3\phi_3}.$$

(54)

$$\xi_{\text{inv}} = \frac{\phi_0}{3\phi_3 - 1} - \frac{2\phi_2}{3\phi_3} + \frac{2\phi_3}{27}.$$

(55)

$$\phi_1 = -\frac{1}{2}\sigma_1^2 EF_T + \frac{1}{2}(1-\gamma)\text{var}(C_T + D_T)$$

$$+ \left(\mu_0 + \mu_1 X - \sigma_2^2 \xi_{\text{fin}}\right) F_T$$

$$+ \left(1-\gamma\right)\xi_{\text{fin}} \text{cov}(A_T + B_T, C_T + D_T)$$

$$+ \left(\mu_0 + \mu_1 X - \sigma_2^2 \xi_{\text{fin}}\right) F_T,$$

(61)

$$\phi_2 = \left(1-\gamma\right)\text{cov}(C_T + D_T)$$

$$+ \left(\mu_0 + \mu_1 X - \sigma_2^2 \xi_{\text{fin}}\right) F_T,$$

(62)

and

$$\phi_3 = \frac{1}{2}(1-\gamma)\sigma_2^2 \text{var}(F_T).$$

(63)

Hence, for any given $\xi_{\text{fin}}$, we can solve the associated $\xi_{\text{inv}}$, which maximizes $U(\gamma)$. Eq. (59), as

$$\xi_{\text{inv}} = -\frac{\phi_2}{4\phi_3} - \frac{\left[\sqrt{q + \frac{4\phi_1^2 - 2\phi_3}{\phi_2}} + \frac{2\phi_3}{27}\right]^3}{2\phi_3},$$

(64)

where

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

(65)

and

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

(66)

If $\xi_{\text{fin}} = \xi_{\text{fin}}^{*1}$ or $\xi_{\text{fin}}^{*2}$ or $\xi_{\text{fin}}^{*3}$, we obtain the corresponding $\xi_{\text{inv}}$ 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 $\xi_{\text{fin}}^{*1}$ and $\xi_{\text{fin}}^{*2}$, 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

\[ dS_t = (\mu_0 + \mu_1 \hat{X}_t + \beta X_t) dt + \sigma_s dB_t, \]

\[ dX_t = (\theta_0 + \theta_1 X_t) dt + \sigma_s dZ_t, \]

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

\[ d\beta = \lambda (\beta - \beta_t) dt + \sigma_\beta dZ^{\beta}_t, \]

where the parameters of this process, i.e., the long-term mean \( \beta \) and reversion speed \( \lambda \), are known to investors. But the investor does not observe the innovation process \( Z^{\beta}_t \) directly and has to infer the realization of \( \beta \) through observations on \( S_t \) and \( X_t \). To complete the model, assume \( E(db_t, dZ^{\beta}_t) = \rho_{\beta t} dt, E(dZ_t, dZ^{\beta}_t) = \rho_{\beta t} dt, \) and \( E(db_t, dZ_t) = \rho dt. \)

Let \( \mathcal{F}_t \) be the investor's filtration. Adapted to \( \mathcal{F}_t \), the least square estimate of \( \beta \) is Gaussian, with mean and variance

\[ b_t = E[\beta_t | \mathcal{F}_t], \]

and

\[ v_t = E[(\beta_t - b_t)^2 | \mathcal{F}_t]. \]

Starting from a Gaussian prior for \( \beta \) with mean \( b_0 \) and variance \( \nu_0 \), the Bayesian updating rule for the conditional mean and variance, \( b_t \) and \( v_t \), are (see, Xia, 2001)

\[ db_t = \lambda (b_t - b_0) dt + \nu_1 d\hat{B}_t + \nu_2 d\hat{Z}_t, \]

\[ dv_t = -2 \lambda \nu_1 + \sigma_\beta^2 - (\nu_1^2 + \nu_2^2 + 2 \nu_1 \nu_2 \rho), \]

where

\[ \hat{B} = \beta, \]

\[ \nu_1 = \frac{\nu_1 (X_t - \hat{X}_t) + \sigma_\beta \sigma_s (\rho_{\beta t} - \rho_{\beta t}))}{\sigma_s (1 - \rho^2)}, \]

\[ \nu_2 = \frac{-\nu_1 (X_t - \hat{X}_t) \rho_{\beta t} + \sigma_\beta \sigma_s (\rho_{\beta t} - \rho_{\beta t}))}{\sigma_s (1 - \rho^2)}, \]

\[ db_t = dB_t + (X_t - \hat{X}_t) (b_t - b_0) dt, \]

and

\[ d\hat{Z}_t = d\hat{Z}_t. \]

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

\[ \xi^*_{\text{opt}} = \frac{\mu_t + \beta_t (X_t - \hat{X}_t) - r}{\sigma^2_t}. \]

Hence, the optimal log-utility level is

\[ U^*_{\text{opt}} = E \log W_T = \int_0^T \left[ r + \xi^*_{\text{opt}} (\mu_t + \mu_t \hat{X}_t + \beta_t (X_t - \hat{X}_t) - r) \right] dt + \log W_t. \]

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

\[ \xi^*_{\text{fix}} = \frac{\mu_t - r + C_T}{\sigma^2_t}, \]

where

\[ C_T = \frac{1}{T} \int_0^T E[\beta \hat{X}_t] dt = \frac{\sigma^2_s \sigma \sigma_s}{(\theta_1 - \lambda)^2} \left[ \frac{e^{(\theta_1 - \lambda)^2 T} - 1}{T} \right]. \]

Intuitively, \( C_T \) captures the covariance between the predictability parameter \( \beta \) and state variable \( X_t \).

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

1. The optimal dynamic learning rule \( \xi^*_{\text{opt}} \) as given by Eq. (79).
2. The optimal fixed strategy \( \xi^*_{\text{fix}} \) as given by Eq. (81).
3. The GMA rule, a combination of coefficients $b_2 (1 - b_2)$ and $C_{mv}$, is denoted as $\text{Fixl} + \text{MA}$, 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, knowledgeable 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

$$ R_t = \mu_t + \sigma_t \eta_t, \quad \text{Eq. (85)} $$

where $\mu_t$ and $\sigma_t$ are defined similarly in Eqs. (27) and (28) with the unknown $\mu_t$, now replaced by the long-term mean $\bar{\mu}$.

The fixed and GMA rules are denoted as $\text{Fixl}$ and $\text{Fixl} + \text{MA}$, respectively, because they are the corresponding strategies of the complete information case.

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

\[ \sigma^2 > \frac{\mu^2 \sigma^2}{\theta_1^2} - 2\mu_{1}\sigma_{1}\sigma_{2}\rho. \]  

(98)

This is because \( \sigma_{2} \) is much smaller relative to \( \sigma_{1} \), and because the correlation \( \rho \) is close to zero. This relation holds for all three calibrated models provided later. Using the unit-free variable \( x = \sqrt{\theta_{1}} \), we can approximate Eqs. (29)-(31) by

\[ C(x) \approx C_1 \left( 1 - \frac{1 - e^{-x^2}}{x^2} \right). \]  

(99)

\[ C_{2,1} \approx C_2 \left( 1 - \frac{1 - e^{-x^2}}{x^2} \right) \]  

and

\[ C_{3} = \frac{\mu_{1} - \sigma_{1}/2}{2\theta_1}. \]  

(102)

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

\[ b_1 \approx C_4 \left( 1 - \frac{1 - e^{-x^2}}{x^2} \right) f(Ax) = C_4 h(x)f(Ax), \]  

(105)

\[ b_2 \approx N(Ax), \]  

(106)

where

\[ A = \frac{C_1}{\sqrt{C_2}} = \frac{\sqrt{3}}{2} \frac{\mu_{1} - \sigma_{1}/2}{\sigma_{1}/\theta_1}. \]  

(107)

\[ C_4 = \frac{C_1}{\sqrt{C_2}}, \]  

(108)

\[ h(x) = \frac{1}{x} \left( 1 - \frac{1 - e^{-x^2}}{x^2} \right). \]  

(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, G, γ), if the investment horizon \( T \) is long enough, then the optimal lag \( L_{opt} \) under the log-utility is approximately given by

\[ L_{opt} \approx \left| \theta_1 \right| \left( 1 + \frac{A^2}{2} + \frac{1}{\sqrt{2}} \left( 1 + \frac{A^2}{2} \right)^{1/2} - \left( \frac{5}{12} + \frac{A^2}{3} \right)^{1/2} \right)^{-1}, \]  

(110)

where \( A = A/\sqrt{2} \) 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 \( \mu_{1} \), stock volatility \( \sigma_{1} \), and state variable mean reversion speed \( |\theta_1| \) given that \( T \) is large. Because \( \mu_{1} \) and \( \sigma_{1} \) are stable across different models, \( L_{opt} \) is mainly driven by differences in \( \theta_1 \).

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 under-performs \( \zeta_{\text{fix}} \) 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 \( \theta_1 \). An alternative initial value for the pure MA is zero. In this case, it can be shown that

\[ L_{opt} \approx \frac{2 \log(\left| \theta_1 \right| T)}{A \left| \theta_1 \right|} \]  

(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 Ait-Sahalia and Brandt (2001) and Brandt and Santa-Clara (2006). 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

---

See Section A.6 of Appendix A for more discussion and for the implementation details.
predictive variable in the model. With the calibrated model and with setting $\gamma = 2$ and $\tau = 5\%$, 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 $U^*_{opt}(W_0)$ be the expected utility based on the optimal dynamic strategy and $U^*_0(W_0)$ be the expected utility based on any of the suboptimal trading strategies, say a fixed strategy. Because $U^*_{opt}(W_0) > U^*_0(W_0)$, there exists $\sigma^* > 0$ such that

$$U^*_{opt}(W_0 - CE) = U^*_0(W_0). \quad (112)$$

$CE$ 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 Brock, Lakonishok, and LeBaron (1992), 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.

### Table 1
List of various portfolio strategies and their comparisons.

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(complete information)</td>
<td>(parameter uncertainty)</td>
<td>(model uncertainty)</td>
</tr>
<tr>
<td>Fix1</td>
<td>Fix1</td>
<td>Fix1</td>
</tr>
<tr>
<td>Fix2</td>
<td>Fix1 + MA</td>
<td>Fix1 + MA</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PureMA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear rule</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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. 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 cutoff 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.

### Table 2
Calibrated model parameters.

The table reports parameter estimates for the following cum-dividend price process:

$$dX_t = (\mu_0 + \mu_1 X_t) dt + \sigma_t dW_t,$$

where $\mu_0, \mu_1, \sigma_t, \beta_t$, and $\sigma_t$ are parameters. $X_t$ is a predictive variable, and $\beta_t$ and $Z_t$ are standard Brownian motions with correlation coefficient $\rho$.

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.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Dividend yield</th>
<th>Term-spread</th>
<th>Payout ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>0.031</td>
<td>0.097</td>
<td>0.282</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>2.072</td>
<td>1.206</td>
<td>-0.292</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.195</td>
<td>0.195</td>
<td>0.194</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.001</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.253</td>
<td>-0.527</td>
<td>-0.027</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.012</td>
<td>0.013</td>
<td>0.050</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.073</td>
<td>0.001</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

---

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

12 The results when $L = 100$ are similar and omitted for brevity.
Table 3
Utility losses versus optimal strategy (L = 50).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Dividend yield</th>
<th>Term-spread</th>
<th>Payout ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>T = 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>8.845</td>
<td>3.895</td>
<td>20.856</td>
</tr>
<tr>
<td>Fix2</td>
<td>7.904</td>
<td>1.568</td>
<td>18.061</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>8.177</td>
<td>2.615</td>
<td>18.639</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>7.895</td>
<td>1.550</td>
<td>18.061</td>
</tr>
<tr>
<td>Pure MA</td>
<td>16.303</td>
<td>13.088</td>
<td>27.692</td>
</tr>
<tr>
<td>MA1</td>
<td>17.762</td>
<td>14.037</td>
<td>28.096</td>
</tr>
<tr>
<td>MA2</td>
<td>17.234</td>
<td>14.423</td>
<td>30.815</td>
</tr>
<tr>
<td>Linear rule</td>
<td>14.801</td>
<td>13.702</td>
<td>33.231</td>
</tr>
<tr>
<td>T = 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>16.680</td>
<td>7.609</td>
<td>31.275</td>
</tr>
<tr>
<td>Fix2</td>
<td>15.171</td>
<td>3.112</td>
<td>30.682</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>15.344</td>
<td>4.646</td>
<td>30.619</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>15.161</td>
<td>3.059</td>
<td>30.681</td>
</tr>
<tr>
<td>Pure MA</td>
<td>29.326</td>
<td>23.931</td>
<td>41.309</td>
</tr>
<tr>
<td>MA1</td>
<td>31.572</td>
<td>25.652</td>
<td>43.109</td>
</tr>
<tr>
<td>MA2</td>
<td>30.694</td>
<td>26.525</td>
<td>43.974</td>
</tr>
<tr>
<td>MA3</td>
<td>30.040</td>
<td>24.874</td>
<td>42.992</td>
</tr>
<tr>
<td>Linear rule</td>
<td>33.409</td>
<td>30.891</td>
<td>58.311</td>
</tr>
<tr>
<td>T = 40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>30.369</td>
<td>14.613</td>
<td>50.694</td>
</tr>
<tr>
<td>Fix2</td>
<td>28.027</td>
<td>6.129</td>
<td>49.495</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>27.907</td>
<td>7.949</td>
<td>50.272</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>27.985</td>
<td>5.880</td>
<td>49.495</td>
</tr>
<tr>
<td>Pure MA</td>
<td>50.036</td>
<td>42.910</td>
<td>59.359</td>
</tr>
<tr>
<td>MA1</td>
<td>53.684</td>
<td>45.284</td>
<td>63.832</td>
</tr>
<tr>
<td>MA2</td>
<td>51.972</td>
<td>45.561</td>
<td>60.039</td>
</tr>
<tr>
<td>MA3</td>
<td>51.136</td>
<td>43.758</td>
<td>61.304</td>
</tr>
<tr>
<td>Linear rule</td>
<td>61.541</td>
<td>58.711</td>
<td>80.391</td>
</tr>
</tbody>
</table>

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 underperformance 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 lengths 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 lengths. 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.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Dividend yield</th>
<th>Term-spread</th>
<th>Payout ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>8.845</td>
<td>3.895</td>
<td>20.856</td>
</tr>
<tr>
<td>Fix2</td>
<td>7.904</td>
<td>1.568</td>
<td>18.061</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>8.091</td>
<td>2.514</td>
<td>18.153</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>7.901</td>
<td>1.547</td>
<td>18.070</td>
</tr>
<tr>
<td>PureMA</td>
<td>15.074</td>
<td>11.668</td>
<td>24.951</td>
</tr>
<tr>
<td>$T = 20$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>16.680</td>
<td>7.609</td>
<td>31.275</td>
</tr>
<tr>
<td>Fix2</td>
<td>15.171</td>
<td>3.112</td>
<td>30.842</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>15.036</td>
<td>4.391</td>
<td>30.531</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>15.166</td>
<td>3.053</td>
<td>30.719</td>
</tr>
<tr>
<td>PureMA</td>
<td>26.873</td>
<td>21.355</td>
<td>38.917</td>
</tr>
<tr>
<td>$T = 40$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>30.369</td>
<td>14.613</td>
<td>50.694</td>
</tr>
<tr>
<td>Fix2</td>
<td>28.027</td>
<td>6.129</td>
<td>49.495</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>27.378</td>
<td>6.697</td>
<td>50.660</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>28.017</td>
<td>5.963</td>
<td>49.641</td>
</tr>
<tr>
<td>PureMA</td>
<td>45.715</td>
<td>36.287</td>
<td>55.856</td>
</tr>
</tbody>
</table>

### 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.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Dividend yield</th>
<th>Term-spread</th>
<th>Payout ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>8.845</td>
<td>3.895</td>
<td>20.856</td>
</tr>
<tr>
<td>Fix2</td>
<td>7.904</td>
<td>1.568</td>
<td>18.061</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>8.153</td>
<td>2.737</td>
<td>18.149</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>7.898</td>
<td>1.521</td>
<td>18.054</td>
</tr>
<tr>
<td>PureMA</td>
<td>16.085</td>
<td>13.215</td>
<td>25.741</td>
</tr>
<tr>
<td>$T = 20$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>16.680</td>
<td>7.609</td>
<td>31.275</td>
</tr>
<tr>
<td>Fix2</td>
<td>15.171</td>
<td>3.112</td>
<td>30.842</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>15.164</td>
<td>4.517</td>
<td>30.559</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>15.162</td>
<td>3.058</td>
<td>30.728</td>
</tr>
<tr>
<td>PureMA</td>
<td>28.607</td>
<td>23.031</td>
<td>40.078</td>
</tr>
<tr>
<td>$T = 40$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix1</td>
<td>30.369</td>
<td>14.613</td>
<td>50.694</td>
</tr>
<tr>
<td>Fix2</td>
<td>28.027</td>
<td>6.129</td>
<td>49.495</td>
</tr>
<tr>
<td>Fix1 + MA</td>
<td>27.379</td>
<td>6.735</td>
<td>50.820</td>
</tr>
<tr>
<td>Fix2 + MA</td>
<td>28.028</td>
<td>6.023</td>
<td>49.683</td>
</tr>
<tr>
<td>PureMA</td>
<td>47.172</td>
<td>38.688</td>
<td>56.718</td>
</tr>
</tbody>
</table>
Statistics | Fix1 | Fix2 | Fix1 + MA | Fix2 + MA | PureMA
--- | --- | --- | --- | --- | ---

### T = 10

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annualized mean</td>
<td>8.703</td>
<td>8.563</td>
<td>8.578</td>
<td>8.531</td>
</tr>
<tr>
<td>Annualized median</td>
<td>8.708</td>
<td>8.568</td>
<td>8.657</td>
<td>8.539</td>
</tr>
<tr>
<td>Annualized SD</td>
<td>5.386</td>
<td>5.081</td>
<td>5.102</td>
<td>5.010</td>
</tr>
<tr>
<td>Annualized Sharpe</td>
<td>0.688</td>
<td>0.701</td>
<td>0.701</td>
<td>0.705</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.008</td>
<td>-0.008</td>
<td>-0.057</td>
<td>-0.022</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.950</td>
<td>2.950</td>
<td>2.967</td>
<td>2.956</td>
</tr>
<tr>
<td>Max DD</td>
<td>31.349</td>
<td>29.511</td>
<td>30.181</td>
<td>29.217</td>
</tr>
</tbody>
</table>

### 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 | 0.994 | 0.995 | 0.903 |
| 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

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annualized mean</td>
<td>8.825</td>
<td>8.619</td>
<td>8.682</td>
</tr>
<tr>
<td>Annualized median</td>
<td>8.872</td>
<td>8.665</td>
<td>8.739</td>
</tr>
<tr>
<td>Annualized SD</td>
<td>2.889</td>
<td>2.663</td>
<td>2.737</td>
</tr>
<tr>
<td>Annualized Sharpe</td>
<td>1.324</td>
<td>1.359</td>
<td>1.345</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.016</td>
<td>-0.016</td>
<td>-0.046</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.009</td>
<td>3.009</td>
<td>2.999</td>
</tr>
<tr>
<td>Max DD</td>
<td>42.718</td>
<td>39.299</td>
<td>41.349</td>
</tr>
</tbody>
</table>

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, 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.

### Table 8

Comparison under parameter uncertainty (T = 10).

| | | | | |
|---|---|---|---|
| \( \sqrt{\psi} \) | \( U_{opt} \) | \( U_{fix1} \) | \( U_{fix1+MA} \) | \( CE_{fix1} \) | \( CE_{fix1+MA} \) |
| \( b_0 = 0 \) | | | | |
| 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 |

PureMA are much smaller. 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 \( \rho_{\beta} \) to be zero. Then, neither Fix1 nor Fix1 + MA depends on the unknown parameter \( \beta \), and \( \psi_{fix1+MA} \) reduces to the optimal fixed rule \( \psi_{fix1} \). In addition, for the mean-reverting process on \( \beta \), we assume \( \beta_{t} \) starts from its calibrated long-term mean, \( \beta_{0} = 2.072 \), 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, Fix\(l\) and Fix\(l + MA\), respectively. The last two columns are the CE or utility losses (in percentage points) of the Fix\(l\) and Fix\(l + MA\) relative to the optimal learning one. Because \(\rho_{\mu} = 0\), the performances of both Fix\(l\) and Fix\(l + MA\) are independent of priors on \(\beta\). The performance of the optimal updating rule depends on the prior. When the prior mean \(b_0 = 0\), both Fix\(l\) and Fix\(l + MA\) depend on the prior. The last two expected utilities associated with the optimal learning and standard error, the third to the fifth columns are the associated loss is the largest. While it is unclear ex ante whether or not \(\sqrt{\sigma^2} = 2\) is clearly the best one, and hence it is not surprising to see that the under-perform the optimal learning rule substantially, when \(\sqrt{\sigma^2} = 1\). When the prior mean moves further away from 0, which is not too relative to the true \(\sigma^2\) 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 \(\beta_0\), the optimal learning rule can now perform worse than either Fix\(l\) or Fix\(l + MA\) when \(\sqrt{\sigma^2} = 1\). 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, Fix\(l\), and its optimal combination with the moving average (MA) strategies, Fix\(l + MA\), 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 \(\beta\) is captured by a mean-reverting process starting from its long-term level \(\beta_0 = 2.0715\). The standard normal prior on \(\beta_0\) has a prior mean \(b_0\) and standard deviation \(\sqrt{\sigma^2}\).

<table>
<thead>
<tr>
<th>(\sqrt{\sigma^2})</th>
<th>(U_{opt})</th>
<th>(U_{fixl})</th>
<th>(U_{fixl+MA})</th>
<th>(CE_{fixl})</th>
<th>(CE_{fixl+MA})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_0 = 0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.503</td>
<td>0.457</td>
<td>0.460</td>
<td>4.590</td>
<td>4.231</td>
</tr>
<tr>
<td>2</td>
<td>0.504</td>
<td>0.457</td>
<td>0.460</td>
<td>4.581</td>
<td>4.322</td>
</tr>
<tr>
<td>3</td>
<td>0.501</td>
<td>0.457</td>
<td>0.460</td>
<td>4.381</td>
<td>4.021</td>
</tr>
<tr>
<td>4</td>
<td>0.491</td>
<td>0.457</td>
<td>0.460</td>
<td>3.470</td>
<td>3.111</td>
</tr>
<tr>
<td>(b_0 = 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.514</td>
<td>0.457</td>
<td>0.460</td>
<td>5.770</td>
<td>5.411</td>
</tr>
<tr>
<td>2</td>
<td>0.515</td>
<td>0.457</td>
<td>0.460</td>
<td>5.790</td>
<td>5.440</td>
</tr>
<tr>
<td>3</td>
<td>0.511</td>
<td>0.457</td>
<td>0.460</td>
<td>5.399</td>
<td>5.040</td>
</tr>
<tr>
<td>4</td>
<td>0.500</td>
<td>0.457</td>
<td>0.460</td>
<td>4.349</td>
<td>3.990</td>
</tr>
<tr>
<td>(b_0 = 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.404</td>
<td>0.457</td>
<td>0.460</td>
<td>-5.301</td>
<td>-5.660</td>
</tr>
<tr>
<td>2</td>
<td>0.414</td>
<td>0.457</td>
<td>0.460</td>
<td>-4.240</td>
<td>-4.599</td>
</tr>
<tr>
<td>3</td>
<td>0.423</td>
<td>0.457</td>
<td>0.460</td>
<td>-3.409</td>
<td>-3.769</td>
</tr>
<tr>
<td>4</td>
<td>0.424</td>
<td>0.457</td>
<td>0.460</td>
<td>-3.259</td>
<td>-3.518</td>
</tr>
<tr>
<td>(b_0 = 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.319</td>
<td>0.457</td>
<td>0.460</td>
<td>-13.750</td>
<td>-14.100</td>
</tr>
<tr>
<td>2</td>
<td>0.338</td>
<td>0.457</td>
<td>0.460</td>
<td>-11.920</td>
<td>-12.279</td>
</tr>
<tr>
<td>3</td>
<td>0.355</td>
<td>0.457</td>
<td>0.460</td>
<td>-10.149</td>
<td>-10.508</td>
</tr>
<tr>
<td>4</td>
<td>0.366</td>
<td>0.457</td>
<td>0.460</td>
<td>-9.090</td>
<td>-9.450</td>
</tr>
</tbody>
</table>

---

Table 10
Comparison under model uncertainty.

The table reports the utility losses of the estimated Fix\(l\) and Fix\(l + MA\) 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 set to equal to five, 10, and 20 years, respectively.

<table>
<thead>
<tr>
<th>Investment</th>
<th>Fix(l + MA)</th>
<th>Fix(l)</th>
<th>Uncertain models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon</td>
<td>(L = 50)</td>
<td>(L = 200)</td>
<td>Wrong Model 1</td>
</tr>
<tr>
<td>Panel A: Dividend yield</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T = 5)</td>
<td>5.228</td>
<td>5.333</td>
<td>5.616</td>
</tr>
<tr>
<td>(T = 10)</td>
<td>13.558</td>
<td>13.359</td>
<td>13.989</td>
</tr>
<tr>
<td>(T = 20)</td>
<td>28.294</td>
<td>27.848</td>
<td>28.366</td>
</tr>
<tr>
<td>Panel B: Term-spread</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T = 5)</td>
<td>1.361</td>
<td>1.420</td>
<td>1.833</td>
</tr>
<tr>
<td>(T = 10)</td>
<td>3.892</td>
<td>3.652</td>
<td>5.408</td>
</tr>
<tr>
<td>(T = 20)</td>
<td>8.735</td>
<td>8.360</td>
<td>11.264</td>
</tr>
<tr>
<td>Panel C: Payout ratio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T = 5)</td>
<td>3.372</td>
<td>3.642</td>
<td>4.190</td>
</tr>
<tr>
<td>(T = 10)</td>
<td>12.313</td>
<td>12.724</td>
<td>16.431</td>
</tr>
<tr>
<td>(T = 20)</td>
<td>34.936</td>
<td>35.467</td>
<td>40.112</td>
</tr>
</tbody>
</table>

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 to be the true one, 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. Zhuo 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 Fix\(l + MA\) 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%.
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 model 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., $\hat{\xi}_{\text{fix}} = b_0/\hat{\sigma}^2$ with $b_0$ 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 \( \theta_1 \), 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 \( y = 1 \), while \( y \) has a value of 2 in the tables. The smaller the \( y \), 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

\[
\begin{align*}
\{ \text{d}X_t &= (\theta_0 + \theta_1 X_t) \text{d}t + \sigma_t \text{d}Z_t, \\
\{ \text{d}Y_t &= (\mu_0 + \mu_1 X_t - \sigma_t^2/2) \text{d}t + \sigma_t \text{d}B_t,
\end{align*}
\]

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

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 \( t \geq 0 \). Integrating Eq. (113) for \( X_t \), we have

\[
X_t = X_0 e^{\theta_1 t} - \frac{\theta_0}{\theta_1} (1 - e^{\theta_1 t}) + \sigma_t \int_0^t e^{\theta_1 (t-s)} \text{d}Z_s.
\]

It follows that \( X_t \) is normally distributed with mean and covariance

\[
\begin{align*}
\text{EX}_t &= EX_0 e^{\theta_1 t} - \frac{\theta_0}{\theta_1} (1 - e^{\theta_1 t}) \\
\text{cov}(X_t, X_s) &= \left[V(0) - \frac{\sigma_t^2}{2\theta_1} (e^{2\theta_1 (t-s)} - 1) \right] e^{\theta_1 (t+s)},
\end{align*}
\]

respectively, where \( \text{EX}_0 \) 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 \( t \rightarrow +\infty \) in Eqs. (115) and (116), i.e.,

\[
\begin{align*}
\bar{X} &= -\frac{\theta_0}{\theta_1} \\
\bar{V} &= -\frac{\sigma_t^2}{2\theta_1}.
\end{align*}
\]

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

\[
\text{EX}_t = \bar{X} = -\frac{\theta_0}{\theta_1},
\]

(119)

and

\[
\text{cov}(x_t, x_s) = -\frac{\sigma_2^2}{2\theta_1} e^{-\theta_1|t-s|}.
\]

(120)

With initial conditions \( X_{|t=0} = X_0 \), \( y_{|t=0} = y_0 \), we integrate Eq. (113) to obtain

\[
\begin{align*}
X_t &= X_0 e^{	heta_1 t} - \frac{\theta_0}{\theta_1} (1 - e^{\theta_1 t}) + \sigma_2 \int_0^t e^{\theta_1(t-s)} dZ_s, \\
y_t &= y_0 + \int_0^t \left( \mu_0 + \mu_1 X_s - \sigma_2^2/2 \right) ds + \sigma_1 B_t.
\end{align*}
\]

(121)

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

\[
M_t = \frac{1}{2} \int_{t-L}^t y_s ds.
\]

(122)

To derive Eq. (15), we note, under constant holding \( \xi_n \), the wealth process is

\[
\log W_t = \log W_0 + rT + \xi_n(\mu_0 - r - \xi_n \sigma_2^2/2)T
\]

and

\[
\text{cov}(\xi_n, \xi_t) = \int_0^T \int_0^T \xi_n \xi_t ds dt + \sigma_2^2 \int_0^T \int_0^T \theta_1 ds dt.
\]

Then, optimizing over \( \xi_n \) the power-utility

\[
\frac{1}{1-\gamma} \text{E}[\exp((1-\gamma) \log W_T)]
\]

\[
= \frac{1}{1-\gamma} \exp(1-\gamma)(\log W_0 + rT
\]

\[
+ \xi_n \mu_0 - r - \xi_n \sigma_2^2/2)T)
\]

\[
\times \exp\left[ (\xi_n \mu_0 \int_0^T X_s dt + \xi_n \sigma_2 \sigma_1 B_T) (1-\gamma) \right].
\]

(123)

we obtain the solution

\[
\bar{\xi}_n = \frac{(\mu_0 - r) + \xi_n E\left[ \frac{1}{T} \int_0^T X_t dt \right]}{\gamma \sigma_2^2 - (1-\gamma) E[X^2](A+2\mu_1 \sigma_1 B)}.
\]

(124)

where

\[
A = \frac{1}{T} \text{var} \left[ \int_0^T X_t dt \right]
\]

(125)

and

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

(126)

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

\[
A = \int_0^T dt \int_0^T ds \text{cov}(X_t, X_s) = \frac{\sigma_2^2}{2\theta_1} \int_0^T dt \int_0^T ds e^{-\theta_1|t-s|}
\]

\[
= \frac{\sigma_2^2}{\theta_1^2} \left( T + \frac{1}{\theta_1} - 1 \right)
\]

(127)

and

\[
B = \int_0^T (X_t, B_T) dt = \frac{\sigma_2^2}{\theta_1} \left( e^{\theta_1 T} - 1 - T \right).
\]

(128)

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 \( t \leq T \)

\[
\langle X_t, B_T \rangle = \sigma_1 \int_0^t e^{\theta_1(t-s)} dZ_s, B_T
\]

\[
= \sigma_1 \int_0^t e^{\theta_1(t-s)} ds = \frac{\sigma_2^2}{\theta_1} (e^{\theta_1 t} - 1).
\]

(129)

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

\[
E y_t = y_0 + (\mu_0 + \mu_1 X_s - \sigma_2^2/2)t
\]

(130)

(131)

and

\[
EM_t = y_0 + (\mu_0 + \mu_1 X_s - \sigma_2^2/2)(t - T/2)
\]

(132)

when \( T > L \). These results allow us to compute the following second moments for \( t > L \):

\[
\langle X_t, X_{t-L} \rangle = \frac{-\sigma_2^2}{2\theta_1} e^{\theta_1 t},
\]

(133)

\[
\langle y_t, X_{t-L} \rangle = \int_0^t \mu_1 (X_s, X_{t-L}) ds + \sigma_2 \int_0^t e^{\theta_1(t-s)} (dW_s, B_t)
\]

\[
= \int_0^t \mu_1 (X_s, X_{t-L}) ds + \int_0^t (1) \mu_1 (X_s, X_{t-L}) ds
\]

\[
+ \sigma_1 \sigma_2 \int_0^t e^{\theta_1(t-s)} ds = \mu_1 \sigma_2^2 (2 - e^{{\theta_1}(t-L)} - e^{\theta_1 t}) - \sigma_2 \sigma_1 \int_0^t e^{\theta_1(t-s)} ds
\]

(134)

\[
\langle y_t, X_{t-L} \rangle = \mu_1 \sigma_2^2 (2 - e^{{\theta_1}(t-L)} - e^{\theta_1 t}) - \sigma_2 \sigma_1 \int_0^t e^{\theta_1(t-s)} ds
\]

(135)

and

\[
\langle y_t, X_{t-L} \rangle = \mu_1 \sigma_2^2 (2 - e^{{\theta_1}(t-L)} - e^{\theta_1 t}) - \sigma_2 \sigma_1 \int_0^t e^{\theta_1(t-s)} ds
\]

(136)

where we have used the fact \( \langle X_s, B_t \rangle = \sigma_1 \int_0^t e^{\theta_1(t-s)} \rho \ ds, \) for \( s \leq t \), an equality

\[
\int_0^t \int_0^t \langle X_s, X_u \rangle ds du = \frac{\sigma_2^2}{\theta_1^2} t + \frac{\sigma_2^2}{\theta_1^2} (1 - e^{\theta_1 t})
\]

(137)
and another equality
\[
(y_t, y_{t-1}) = (y_{t-L}, y_{t-L}) + \int_{t-L}^t \mu_t(y_t, y_{t-1}) \, ds
\]
\[
= \left( \sigma_t^2 + \frac{(\mu_t \sigma_t)^2}{\theta_t^2} - \frac{2 \mu_t \sigma_t \rho_t}{\theta_t} \right) (t-L)
+ \left( \frac{(\mu_t \sigma_t)^2}{2 \theta_t^2} - \frac{\mu_t \sigma_t \rho_t}{\theta_t} \right) (1-e^{-\theta_t})
\times (1-e^{\theta_t(t-L)} + e^{\theta_t L} - e^{\theta_t t}),
\]
\( (138) \)

Next, we compute the following second moments involving \( M_t \) using Eqs. (134) and (138):
\[
(X_t, M_t) = \frac{1}{L} \int_{t-L}^{t} \left( X_t, X_t \right) \, ds
\]
\[
= \frac{1}{L} \left( \sigma^2 + \frac{(\mu \sigma)^2}{\theta^2} - \frac{2 \mu \sigma \rho}{\theta} \right) (t-L)
+ \left( \frac{(\mu \sigma)^2}{2 \theta^2} - \frac{\mu \sigma \rho}{\theta} \right) (1-e^{-\theta})
\times \left( \frac{1}{\theta^ L} - e^{\theta(t-L)} + e^{\theta L} \right),
\]
\( (139) \)

and
\[
(Y_t, M_t) = \frac{1}{L} \int_{t-L}^{t} \left( Y_t, Y_t \right) \, ds
\]
\[
= \left( \sigma^2 + \frac{(\mu \sigma)^2}{\theta^2} - \frac{2 \mu \sigma \rho}{\theta} \right) \left( \frac{t-L}{2} \right)
+ \left( \frac{(\mu \sigma)^2}{2 \theta^2} - \frac{\mu \sigma \rho}{\theta} \right) \frac{1}{\theta^L} (1-e^{-\theta})
\times \left( \frac{1}{\theta^ L} - e^{\theta(t-L)} + e^{\theta L} \right),
\]
\( (140) \)

Finally, to compute \( (M_t, M_t) \), we note first
\[
M_t = \frac{1}{L} \int_{t-L}^{t} y_t \, ds = \frac{1}{L} \int_{0}^{t} \left( y_{t-L} + (y_{t-L} - y_{t-1}) \right) \, ds
\]
\[
= y_{t-L-L} + \frac{1}{L} \int_{0}^{t} \tilde{y}_{t-1-L} \, ds. \tag{141}
\]

where \( \tilde{y}_{t-L-L} = y_{t-L-L} - y_{t-1-L} \). Then, we can write \( (M_t, M_t) \) as
\[
(M_t, M_t) = \left( \left( y_{t-L-L} + \frac{1}{L} \int_{0}^{t} \tilde{y}_{t-1-L} \, ds \right), \left( y_{t-L-L} + \frac{1}{L} \int_{0}^{t} \tilde{y}_{t-1-L} \, ds \right) \right)
\]
\[
= (\tilde{M}_t, \tilde{M}_t) + \frac{1}{L} \int_{0}^{t} \left( y_{t-L-L} - y_{t-L-L} \right) \, ds = (\tilde{M}_t, \tilde{M}_t), \tag{142}
\]

where \( \tilde{M}_t = (1/L) \int_{0}^{t} y_t \, ds \). Using Eq. (138), we obtain
\[
(\tilde{M}_t, \tilde{M}_t) = \frac{1}{L} \int_{0}^{t} \left( \sigma^2 + \frac{(\mu \sigma)^2}{\theta^2} - \frac{2 \mu \sigma \rho}{\theta} \right) \, ds
\]
\[
= \left( \frac{1}{L} \right) \left( \sigma^2 + \frac{(\mu \sigma)^2}{\theta^2} - \frac{2 \mu \sigma \rho}{\theta} \right)
+ \left( \frac{(\mu \sigma)^2}{2 \theta^2} - \frac{\mu \sigma \rho}{\theta} \right) \frac{1}{\theta^L} (1-e^{-\theta})
\times \left( \frac{1}{\theta^ L} - e^{\theta(t-L)} + e^{\theta L} \right), \tag{143}
\]

For the term \( \frac{1}{L} \int_{0}^{t} \left( y_{t-L-L} - y_{t-L-L} \right) \, ds \), 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
\[
(M_t, M_t) = \left( \sigma_t^2 + \frac{(\mu_t \sigma_t)^2}{\theta_t^2} - \frac{2 \mu_t \sigma_t \rho_t}{\theta_t} \right) \left( t-L \right)
+ \left( \frac{(\mu_t \sigma_t)^2}{2 \theta_t^2} - \frac{\mu_t \sigma_t \rho_t}{\theta_t} \right) \frac{1}{\theta_t^L} \left( 1 - e^{\theta_t(t-L)} \right)
\times \left( \frac{1}{\theta_t^ L} - e^{\theta_t} + e^{\theta_t L} \right) \left( 1 - e^{\theta_t(t-L)} \right), \tag{144}
\]

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
\[
n_1 = \frac{\theta_0}{\theta_1}, \tag{145}
n_2 = y_0 + \left( \frac{\mu_0 - \theta_0}{\theta_1} - \frac{\sigma_0^2}{2} \right) t, \tag{146}
n_3 = y_0 + \left( \frac{\mu_0 - \theta_0}{\theta_1} - \frac{\sigma_0^2}{2} \right) \left( t - \frac{L}{2} \right), \tag{147}
\]

and covariance matrix \( D = (D_{ij}) \) given by
\[
D_{11} = -\frac{\sigma_1^2}{\theta_1}, \tag{148}
D_{22} = \frac{\sigma_2^2}{\theta_1} + \frac{2 \mu_0 \sigma_0 \rho_0}{\theta_1} t
+ \left( \frac{\mu_0 \sigma_0^2}{\theta_1} - \frac{\sigma_0 \rho_0}{\theta_1} \right) \left( 1 - e^{\theta_1 t} \right), \tag{149}
D_{23} = \frac{\sigma_2^2}{\theta_1} + \frac{2 \mu_0 \sigma_0 \rho_0}{\theta_1} \left( t - \frac{L}{2} \right)
+ \left( \frac{\mu_0 \sigma_0^2}{\theta_1} - \frac{\sigma_0 \rho_0}{\theta_1} \right) \left( 1 - e^{\theta_1 t} \right)
\times \left( 1 - e^{\theta_1(t-L)} - \frac{2}{\theta_1 L} \right) \left( 1 - e^{\theta_1 L} \right) \left( 1 - e^{\theta_1(t-L)} \right), \tag{150}
D_{12} = \frac{\mu_0 \sigma_0^2}{\theta_1} \left( 1 - e^{\theta_1 t} \right), \tag{151}
D_{13} = \frac{\mu_0 \sigma_0^2}{\theta_1} \left( 1 - e^{\theta_1 t} \right)
\times \left( 1 - e^{\theta_1 L} \right) \left( 1 - e^{\theta_1(t-L)} \right), \tag{152}
D_{23} = \frac{\mu_0 \sigma_0^2}{\theta_1} \left( 1 - e^{\theta_1 L} \right)
\times \left( 1 - e^{\theta_1(t-L)} \right), \tag{153}
D_{33} = \frac{\sigma_3^2}{\theta_1} + \frac{2 \mu_0 \sigma_0 \rho_0}{\theta_1} \left( t - \frac{L}{2} \right)
+ \left( \frac{\mu_0 \sigma_0^2}{\theta_1} - \frac{\sigma_0 \rho_0}{\theta_1} \right) \left( 1 - e^{\theta_1 t} \right)
\times \left( 1 - e^{\theta_1(t-L)} \right), \tag{154}
\]
With Lemma 1, the proof of Eqs. (27) and (28) follows from Lemma 2.

**Lemma 2.** Let $X_t = X_t - \bar{X}$ and $Z_t = Y_t - M_t$. Then $(X_t, Z_t)$ is normally distributed with mean $m = (m_1, m_2 - m_3)$, and covariance $C = (C_{ij})$ given by

\[
C_{11} = D_{11},
\]

\[
C_{22} = D_{22} + D_{33} - 2D_{23},
\]

and

\[
C_{12} = D_{12} - D_{13}.
\]

Moreover,

\[
E[X_t, Z_t] = m_t N\left(\frac{m_2}{\sigma_2}, \frac{C_{12}}{\sigma_2}, \frac{-m_2}{\sigma_2}\right).
\]

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 $(\sigma_x, \sigma_z)$, and correlation $\rho$, i.e.,

\[
E[x_1, z_1] = m_t N\left(\frac{m_2}{\sigma_z}, \frac{C_{12}}{\sigma_z}, \frac{-m_2}{\sigma_z}\right).
\]

After standardization, we can write

\[
\tilde{x} = \frac{x - m_x}{\sigma_x}, \quad \text{and} \quad \tilde{z} = \frac{z - m_z}{\sigma_z},
\]

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

\[
E[x_1, z_1] = m_t N\left(\frac{m_2}{\sigma_z}, \frac{C_{12}}{\sigma_z}, \frac{-m_2}{\sigma_z}\right).
\]

This implies that

\[
E[x_1, z_1] = m_t N\left(\frac{m_2}{\sigma_z}, \frac{C_{12}}{\sigma_z}, \frac{-m_2}{\sigma_z}\right).
\]

which proves Eq. (159). □
Because the value function for log-utility associated with $\xi_{\text{ext}}$ is

$$U_{\text{ext}} = \log W_0 + r T + \frac{(u_0 + u_1 \hat{X} - r)^2}{2 \sigma^2},$$

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 $\xi_{\text{fix}} = \xi_{\text{mv}}$ and optimize $U_{\text{GMA}}(\xi_{\text{fix}}, \xi_{\text{mv}})$ over $\xi_{\text{mv}}$ alone. Similar algebra yields the solution. The proof of Proposition 3 follows analogously.

A.3. Proof of Eq. (64)

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

$$\max f(\xi_{\text{mv}}) = \xi_{\text{mv}}(\phi_0 + \phi_1 \xi_{\text{mv}} + \phi_2 \xi_{\text{mv}}^2 + \phi_3 \xi_{\text{mv}}^3).$$

The first-order condition is

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

which in turn can be transformed to

$$y^3 + py + q = 0,$$

where

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

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

$$q^2 + 4p^3 > 0.$$  

(178)

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

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

(179)

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

$$\xi_{\text{mv}} = -\frac{\phi_2}{4 \phi_3} + y^*.$$  

(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 $U_{\text{GMA}} - U_{\text{fix}}$ and $U_{\text{GMA}} - U_{\text{fix}}$, and ignore some constants, the target functions become

$$U_1 = \frac{b_1}{b_2(1 - b_2)} \left(1 - \frac{L}{\hat{T}}\right) = V_1 \left(1 - \frac{L}{\hat{T}}\right)$$  

(181)

and

$$U_2 = \frac{b_1^2}{b_2^2} \left(1 - \frac{L}{\hat{T}}\right) = \frac{b_1^2}{b_2^2} \left(1 - \frac{L}{\hat{T}}\right).$$  

(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

$$V_2 = \frac{2b_1^2 b_2 - b_1 b_2^2}{b_2^2} = 0.$$  

(183)

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

$$2h(x)f(\hat{X}) - 2A \hat{x} h(x)f(\hat{X}) - \frac{Ah(x)f^2(\hat{X})}{N(\hat{X})} = 0.$$  

(184)

This is a transcendental 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

$$b_1 = h(x) \cdot f(\hat{X}).$$  

(185)

The Taylor expansion for $h(x)$ is

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

(186)

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

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

(187)

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

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

(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 $N(\hat{X})$ and $N(-\hat{X})$, and hence

$$V_1 \approx \frac{h^2(x)f^2(\hat{X})}{N(\hat{X})N(-\hat{X})} = \frac{1}{C} \cdot N(\hat{X}) [h(x) \sqrt{f(\hat{X})}].$$  

(189)

where we have used the approximation $N(-\hat{X}) \approx C \cdot f(\hat{X})$ for $\hat{X} > 0$ and large. Similar to the earlier case, we can ignore $N(\hat{X})$, and hence the target function becomes $h(x) \sqrt{f(\hat{X})}$. This has the same form as Eq. (185) with $A^2/2$ plays the role of earlier $A$. Therefore, the solution follows.
Finally, to derive Eq. (111), we need to maximize $U_3 = \frac{\mu^T S - \sigma^2}{2\sigma^2}$. Similarly, this can be replaced by a target function

$$V_3 = [\mu_c h(xf(Ax) + C_0 N(Ax))] \cdot (1 - x^2)$$

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

$$\frac{\partial V_3}{\partial x} = 0.$$  

Because there is only one solution, we can verify that

$$\frac{\partial V_3}{\partial x} = 0.$$  

and hence we can reduce the first-order condition to

$$\frac{\partial V_3}{\partial x} = 0.$$  

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

$$L(\Theta) = \sum_{t=1}^T \log(f(R_t, X_t) \mid X_{t-1}; \Theta)$$

where $\Theta = (a_{11}, a_{12}, b_{11}, b_{12}, \sigma_1, \sigma_2, \sigma_3)$ with

$$a_{11} = \left(\mu_0 - \frac{1}{2} \sigma_1^2 - \frac{\mu^T h_0}{\theta_1} \frac{\partial^2 f}{\partial \theta^2} \right) \Delta t + \frac{\mu \theta_0}{\theta_1} (e^{\sigma_1 \Delta t} - 1),$$

$$a_{12} = \frac{\mu^T h_1}{\theta_1} (e^{\sigma_1 \Delta t} - 1),$$

$$b_{11} = \frac{\partial^2 f}{\partial \theta^2} \frac{\partial^2 f}{\partial \theta^2} \Delta t + \frac{\mu^T h_0}{\theta_1} \frac{\partial^2 f}{\partial \theta^2} \Delta t + \frac{1}{2 \theta_1} (e^{2\sigma_1 \Delta t} - 1) \frac{\sigma_1^2}{\theta_1} + \frac{\mu^T h_0}{\theta_1} (e^{2\sigma_1 \Delta t} - 1),$$

$$b_{12} = e^{\sigma_1 \Delta t}.$$  

A.6. The linear rule

Ait-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:

$$\xi_j = \zeta_{j,t+1} + \zeta_{j,t+1} X_{t+1}, \quad j = 1, \ldots, H,$$

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 $R_f = 1 + r_t$ 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

$$\tilde{R}_{t+1} = [R_f^{t+1} - R_f^{t+1}]_{j=0}^{t=H-1}, \quad [R_f^{t+1} X_{t+1}]_{j=0}^{t=H-1},$$

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

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

$$\max_{\theta_t} U_t[R_f^{t+1} + \theta_t \tilde{R}_{t+1}].$$

where $\theta_t$, $1 \times 2H$, 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.,

\[
E_t(u(W_{t+1}) = E_t(u(W_{t+1}^R)) + \left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)^2
\]

\[
+ \left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)^3
\]

\[
+ \frac{1}{2}\left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)^4. \quad (208)
\]

As a result,

\[
\theta_t \approx - E_t\left[u\left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)\right]^{-1}
\]

\[
\times E_t\left[u\left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)\right] W_t
\]

\[
+ E_t\left[u\left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)\right]W_t^2
\]

\[
+ E_t\left[u\left( W_tR_t^F(W_t|\hat{t}_t, t_{t-1})\right)\right]W_t^3. \quad (209)
\]

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


