Asset pricing and portfolio choice with technical analysis

Author:
Kwong, Tsz Wang

Publication Date:
2017

DOI:
https://doi.org/10.26190/unsworks/19450

Asset Pricing and Portfolio Choice with Technical Analysis

Tsz Wang Kwong

The School of Banking and Finance

The University of New South Wales

A dissertation submitted to the University of New South Wales in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

7 March 2017

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Kwong

First name: Tsz Wang

Other name/s:

Abbreviation for degree as given in the University calendar: PhD

School: Banking and Finance

Faculty: Business

Title: Asset Pricing and Portfolio Choice with Technical Analysis

Abstract 350 words maximum: (PLEASE TYPE)

Technical analysis is the study of market movements, primarily through the use of past prices and volumes, for the purpose of forecasting future price trends. Despite its popularity among practitioners, academics tend to be skeptical about its true usefulness. One of the major reasons is that it lacks a theoretical basis in finance theory. Although there is increasing empirical evidence in favor of its effectiveness, the empirical debate remains unsettled, meanwhile the progress on strengthening its theoretical basis is relatively slow. To understand better technical analysis as an important and popular investment tool, this thesis aims to further tie technical analysis to modern finance theory in an attempt to tighten this gap in the literature. This thesis includes two chapters that study portfolio choice problems and two additional chapters that study asset pricing problems, in which investors make strategic use of information from technical analysis, specifically the moving averages. Our model approach provides several new insights to the field. We develop a model to examine the effects of the uncertain predictive power of moving averages on portfolio choice. We find that investors accounting for such uncertainty allocate substantially less wealth to stocks and are more conservative in market timing for longer horizons. Furthermore, the utility loss of ignoring this uncertainty can be sizable and increases with horizon at an increasing rate. We present another portfolio choice model to theoretically illustrate that moving averages can be useful for investment when stock returns are correlated. We also formulate an asset pricing model and propose some plausible equilibria in which future prices can be predicted by moving averages. This model provides a theoretical basis for some recent empirical findings that moving averages have predictive power. We further formulate a similar asset pricing model which emphasizes development of estimation and testing strategies to empirically test the proposed equilibria. Using S&P 500 index and dividend data for the period January 1871 to December 2015, we empirically reject the possibility that investors' trend following behavior is the driver of the stock market in the long run.

Declaration relating to disposition of project thesis/dissertation

7 March 2017
Date

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

Abstract

Technical analysis is the study of market movements, primarily through the use of past prices and volumes, for the purpose of forecasting future price trends. Despite its popularity among practitioners, academics tend to be skeptical about its true usefulness. One of the major reasons is that it lacks a theoretical basis in finance theory. Although there is increasing empirical evidence in favor of its effectiveness, the empirical debate remains unsettled, meanwhile the progress on strengthening its theoretical basis is relatively slow. To understand better technical analysis as an important and popular investment tool, this thesis aims to further tie technical analysis to modern finance theory in an attempt to tighten this gap in the literature. This thesis includes two chapters that study portfolio choice problems and two additional chapters that study asset pricing problems, in which investors make strategic use of information from technical analysis, specifically the moving averages. Our model approach provides several new insights to the field. We develop a model to examine the effects of the uncertain predictive power of moving averages on portfolio choice. We find that investors accounting for such uncertainty allocate substantially less wealth to stocks and are more conservative in market timing for longer horizons. Furthermore, the utility loss of ignoring this uncertainty can be sizable and increases with horizon at an increasing rate. We present another portfolio choice model to theoretically illustrate that moving averages can be useful for investment when stock returns are correlated.

We also formulate an asset pricing model and propose some plausible equilibria in which future prices can be predicted by moving averages. This model provides a theoretical basis for some recent empirical findings that moving averages have predictive power. We further formulate a similar asset pricing model which emphasizes development of estimation and testing strategies to empirically test the proposed equilibria. Using S&P 500 index and dividend data for the period January 1871 to December 2015, we empirically reject the possibility that investors' trend following behavior is the driver of the stock market in the long run.

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of the thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed: Tsz Wang Kwong

Date: 7 March 2017

Copyright and DAI Statement

Authenticity Statement

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signature: Tsz Wang Kwong

Date: 7 March 2017

Acknowledgements

I would like to thank all those people who made this thesis possible and an unforgettable experience for me.

First of all, I would like to express my gratitude to my three supervisors, Prof David Feldman, Dr David Colwell, and Dr Christopher Gibbs, who offered their continuous advice and encouragement throughout the course of this thesis.

I am particularly thankful to my parents for their love and patience. I next wish to thank my sworn brother for his emotional support in my tough times. I am so blessed to have them in my life.

Finally, my very sincere thanks to my friends and colleagues for making my life colorful during my PhD study. Thank you indeed for all the wonderful memories you made for me.

Contents

Chapter 1

Introduction

1.1 Introduction

Technical analysis is the study of market movements, primarily through the use of past prices and volumes, for the purpose of forecasting future price trends. Several techniques and tools in technical analysis, most notably moving averages and charting, have been widely used by practitioners. Murphy (1999) summarizes that there are three rationales for technical analysis. First, it is believed that price actions reflect shifts in demand and supply. As a result, prices should contain important information about future movements of that market. Second, it is also believed that prices move in trends. In other words, prices are more likely to move in the same direction until they reverse. Therefore, trying to detect the beginning of a trend or a price reversal is the primary objective of technical analysis. Third, technical investors believe that history repeats itself that many price chart patterns are results of human psychology, which are unlikely to change from generation to generation. Thus, in some sense the future is considered just a repetition of the past.

However, despite the popularity of technical analysis among practitioners, academics tend to be skeptical about its true usefulness. Following Zhu and Zhou (2009), the following three reasons may explain this attitude towards technical analysis. The first reason is that there are limited theoretical studies to justify why technical analysis can have value under certain conditions. While there are theoretical models, notably noisy rational expectations models (see, *e.g.* Working, 1958; Brown and Jennings, 1989; Wang, 1993) and feedback models (see, *e.g.* De Long *et al.*, 1990; Shleifer and Summers, 1990), showing past prices or volumes are useful for forecasting future prices, they are not closely tie to the actual techniques and tools used in technical analysis. There is also a literature on heterogeneous agent models (see, *e.g.* Brock and Hommes, 1997, 1998; Boswijk *et al.* 2007; Hommes, 2013) studying the impacts of different trading or forecasting rules on asset prices, however, the primary objective is to replicate stylized facts of financial time series and thus these studies are empirically oriented.

The second reason is that the efficient market hypothesis used to be so widely accepted that many academics used to posit that share prices should exhibit no serial dependencies, meaning that past prices should be useless to forecast future prices. Although there is now ample evidence that the stock market is indeed predictable, in the sense that future stock returns are correlated with the current values of some observable economic variables, how such predictability allows for the usefulness of technical analysis is not well investigated. The third reason is that empirical findings are mixed and inconclusive. Indeed, it is not uncommon that later studies challenge the statistical validity of previous results. At this stage, it is unlikely that the statistical debate on the usefulness of technical analysis will be settled soon.

1.1.1 Motivation for the Study

Although there is increasing empirical evidence for the usefulness of technical analysis (see, *e.g.* Brock, Lakonishok, and LeBaron, 1992; Han, Yang, and Zhou, 2013; Neely *et al.*, 2014), the progress on strengthening its theoretical basis is relatively slow. In particular, little attention has been paid to the strategic use of information generated by technical analysis on portfolio choice and how asset prices are dynamically affected by investors' use of such information. Previous studies assessing the effectiveness of technical analysis commonly assume that investors use an "all-or-nothing" approach, namely investing 100% of wealth into the stock when the technical trading rule says buy but nothing otherwise. This is a naive use of information from technical analysis in the sense that this would be too risky for risk-averse investors, who would not prefer great fluctuations in wealth. For asset pricing studies, there are only a handful of dynamic models admitting equilibrium prices as a function of technical indicators in closed-form. Moreover, it appears that most asset pricing models are confined to solving for a market equilibrium but fall short in providing a mechanism how the prices adjust to that equilibrium. Considered how relevant and important technical analysis is for real-world investment, there is a need to promote more theoretical studies in the literature.

1.1.2 Aim and Scope

Following the footsteps of some recent attempts, see, *e.g.* Zhu and Zhou (2009) and Zhou and Zhu (2013), this thesis aims to further tie technical analysis to modern finance theory in an attempt to tighten this gap in the literature. Unfortunately, due to the analytical nature of our studies, we have to restrict our attention to the moving averages, which are more mathematically tractable relative to the information visually deduced from charting patterns. Although focusing only on moving averages may appear to be narrow, moving averages are undoubtedly among the most popular and easy-to-use classes of technical trading rules. Moreover, academically, the seminal work by Brock, Lakonishok, and LeBaron (1992) stimulates much subsequent research on the effectiveness of moving averages as an investment tool. Therefore, moving averages are in their own right practically and academically important.

1.1.3 Significance of the Study

This thesis attempts to contribute to the literature in several ways. First, it provides new insights to the field through studying technical analysis in the perspectives of asset pricing and portfolio choice. Second, it demonstrates various modelling and solution techniques in different asset pricing and portfolio choice problems. Third, also as the intended outcome of this thesis, it helps promote building a stronger theoretical basis for the use of technical analysis as an investment tool.

1.1.4 Structure of the Thesis

This thesis consists of five further chapters. While the next four chapters are structured as standalone research papers, they can be conceptually separated into two parts: the next two chapters study portfolio choice problems and the further next two chapters study asset pricing problems. In Chapter 2, we develop a model to examine the effects of the uncertain predictive power of moving averages on portfolio choice. This chapter is characterized by an investor who uses a Bayesian approach to continuously infer useful information from moving-average signals for investment. In Chapter 3, we present another portfolio choice model to theoretically illustrate that moving averages can be useful for investment when stock returns are correlated. We show that moving-average based portfolio strategies are more profitable than strategies ignoring time-varying investment opportunities.

In Chapter 4, we present an asset pricing model to demonstrate the predictive power of moving averages on future stock prices as an equilibrium phenomenon. This model is characterized by a representative investor who formulates forecasts of future stock prices based on different sources of information, one of which is moving averages. Such forecasts then in turn affect equilibrium prices. Different equilibria are therefore obtained by imposing different assumptions on the investor's forecasting rule. We then assess the stability of these equilibria and comment on their plausibility. In Chapter 5, we formulate another asset pricing model based on that of Chapter 4 but with important modifications. Unlike the analytical study in the previous chapter, emphasis is instead put on developing estimation and testing strategies to examine whether the proposed equilibria are empirically supported. The general strategy is that we can make use of a set of orthogonality conditions implied by the equilibrium pricing equation and the investor's forecasting rule for GMM estimation. Using S&P 500 index and dividend data for the period January 1871 to December 2015, we empirically reject the possibility that investors' trend following behavior is the driver of the stock market in the long run. Finally, Chapter 6 summarizes and concludes this thesis.

1.2 References

Boswijk, H. P., Hommes, C. H., & Manzan, S. (2007). Behavioral heterogeneity in stock prices. Journal of Economic Dynamics and Control, 31(6), 1938-1970.

Brock, W. A., & Hommes, C. H. (1997). A rational route to randomness. Econometrica, 65(5), 1059-1095.

Brock, W. A., & Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic dynamics and Control, 22(8), 1235-1274.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

Brown, D. P., & Jennings, R. H. (1989). On technical analysis. Review of Financial Studies, 2(4), 527-551.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. Journal of Political Economy, 98(4), 703-738.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis, 48(5), 1433-1461.

Hommes, C. (2013). Behavioral rationality and heterogeneous expectations in complex economic systems. Cambridge University Press.

Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. Penguin.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60(7), 1772-1791.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. Journal of Economic Perspectives, 4(2), 19-33.

Wang, J. (1993). A model of intertemporal asset prices under asymmetric information. Review of Economic Studies, 60(2), 249-282.

Working, H. (1958). A theory of anticipatory prices. American Economic Review, 48(2), 188-199.

Zhou, G., & Zhu, Y. (2013). An equilibrium model of moving-average predictability and time-series momentum. Unpublished working paper, Washington University in St. Louis.

Zhu, Y., & Zhou, G. (2009). Technical analysis: An asset allocation perspective on the use of moving averages. Journal of Financial Economics, 92(3), 519- 544.

Chapter 2

Technical Analysis with Uncertain Predictive Power: The Effects on Portfolio Choice

Deviating from a conventional statistical testing approach, we analyze the economic relevance of technical analysis. Specifically, we assess how uncertainty in the predictive power of moving average signals affects investors' portfolio choice. Calibrating our model with CRSP index data, we find that investors accounting for such uncertainty allocate substantially less to stocks and are more conservative in market timing for longer horizons. Furthermore, the utility loss of ignoring this uncertainty can be sizable and increases with horizon at an increasing rate.

2.1 Introduction

When an investor receives a "buy" signal from a technical trading rule, how much should he trust this signal and adjust his allocations to stocks? The literature provides limited guidance to answer this question because it is commonly assumed that investors use a naive "all-or-nothing" strategy, namely allocating 100% of wealth to stocks whenever they observe a buy signal but nothing otherwise. This assumption, however, is unrealistic because it overlooks at least two important issues.

First, investors do not necessarily have strong faith in technical analysis and thus betting entire wealth on a buy signal is far too risky. After all, empirical findings of its usefulness are mixed and inconclusive. Early studies are generally skeptical about its usefulness and this perhaps develops a preconception about this investment tool (see, *e.g.*, Famma and Blume, 1966; Jensen and Benington, 1970). Although later studies increasingly provide encouraging evidence (see, *e.g.*, Brock, Lakonishok, and LeBaron, 1992; Lo, Mamaysky, and Wang, 2000; Han, Yang, and Zhou, 2013; Neely *et al.*, 2014), its true effectiveness is still in debate (see, *e.g.*, Allen and Karjalainen, 1999; Ready, 2002; Bajgrowicz and Scaillet, 2012).

Second and more importantly, investors' allocations to stocks should be optimally chosen in a utility maximization framework: an investor's optimal allocation may depend on, for example, his degree of risk aversion, wealth level, prior belief and current assessment of the predictive power of the trading rule. When stock returns are deemed to be predictable, his investment horizon may be also relevant.

Failing to account for these two issues, the *ad hoc* "all-or-nothing" assumption is unlikely to reflect a legitimate use of technical analysis. In this article, we study portfolio choice with technical analysis in a utility maximization framework, with an emphasis on how uncertainty in the predictive power of technical analysis affects portfolio choice. Specifically, we consider an investor who uses a linear prediction model to forecast stock returns with zero-one moving average signals. Uncertainty in predictive power is then formulated as an unknown slope parameter. We assume that the investor uses a simple Bayesian approach[1] to update prior beliefs about the model parameters and then incorporates the precision of the estimates into his expected utility. He then chooses how to allocate his wealth optimally, between a risk-free asset and a risky stock, by maximizing expected utility. This optimization problem is commonly called *portfolio choice with parameter uncertainty*, and the uncertainty in model parameters is called *estimation risk*.

This article contributes to the analysis and understanding of portfolio choice with technical analysis. First, we develop a model to examine the effects of uncertain predictive power of moving average signals on portfolio choice. Second, we derive an approximate solution for the optimal allocation to stocks. Third, we develop a simple numerical procedure to account for and decompose estimation risk. Fourth, calibrating the model with real data, we show that shorter horizon investors (say, no more than five years) bear little utility loss even if they ignore estimation risk. By contrast, such utility loss is sizable for investors with longer horizons. This finding helps explain why short-term speculators appear to use technical analysis with less scepticism.

[1] There are alternative econometric approaches to tackle estimation risk, for example, out-of-sample testing, correction factors, and randomization methods such as bootstrapping and Monte Carlo simulation. However, choosing the most appropriate approach is beyond the scope of this article.

Although there is limited research directly addressing our topic, we do get some insights from the literature. The effects of uncertain return predictability on portfolio choice are studied by, for example, Stambaugh (1999), Barberis (2000), and Xia (2001). These studies illustrate that in general ignoring estimation risk results in a substantial opportunity cost to the investor. We, however, consider technical signals rather than the dividend yields prominently studied in the literature, and we show that it is more mathematically challenging to account for estimation risk in this framework. This is because the law of motion of the predictive variable, namely the technical signals, is implicitly determined by the return prediction model itself (as a function of past stock prices). Therefore, the same source of estimation risk, namely having unknown parameters in the prediction model, *concurrently* affects the forecasts of future returns and technical signals. This recursive structure implies some important differences to that of other studies when the law of motion of dividend yields is independent of the return prediction model.

Zhu and Zhou (2009) develop a model to justify why technical analysis, specifically the moving average trading rule, can provide useful information for portfolio choice. Our model approach differs from theirs in that we explicitly use a prediction model to represent the investor's perceived law of motion for returns and he continuously updates the model parameters to make conditional forecasts. Based on a similar conceptual framework to that of Zhu and Zhou (2009), Zhou, Zhu, and Qiang (2012) test their one-period moving average strategies and find that they outperform other strategies that disregard the usefulness of moving averages. We instead consider an investor who rebalances continuously and shift the focus to study the effects of estimation risk on portfolio choice and investor welfare in the context of technical analysis, which attract little attention in the literature.

As in Zhu and Zhou (2009) and Zhou, Zhu, and Qiang (2012), we do not consider the possible impact on investors' consumption for simplicity. This direction is definitely worth future research. Interested readers may refer to, for example, Brandt *et al.* (2005), Lundtofte (2008), and Xia (2001), for consumption-investment models under parameter uncertainty. We are also aware that the nonparametric Euler condition approach by Brandt (1999) appears to be applicable if one defines a set of finite states based on some technical signals.

Beyond academic studies, technical analysis has been widely used by sophisticated practitioners and is considered an important tool for stock investment. Menkhoff (2010) carries out a survey of fund managers in 2003/2004.[2] The respondents are 692 fund managers in five countries: United States, Switzerland, Germany, Italy, and Thailand and the majority of the surveyed fund managers specialize in stock markets. The survey indicated that 87% of respondents use technical analysis and a major group (18%) prefers technical analysis to other tools for investment decisions, including learning fundamental information from the market. This survey shows that a substantial proportion of technical analysis users has high reasoning power. Therefore, it is relevant to consider a more realistic use of technical analysis in a portfolio choice model and examine its economic relevance by studying the model implications.

The article is organized as follows. The next section discusses some issues about technical analysis and portfolio choice with a prediction model. Section 2.2 introduces our model. Section 2.3 calibrates the model with actual data. Section 2.4 studies the model implications. Section 2.5 summarizes and concludes the article.

[2] The author justifies his sample choice with two reasons: first, fund managers have evolved as the most important group in modern financial markets; second, fund managers are highly qualified market participants compared to typical individual investors.

2.2 Background

There are four issues about technical analysis that are worth clarifying. First, if stock returns are predictable[3], then some technical signals may have predictive power. By contrast, if stock returns are simply random noises, then no technical signal, not even fundamental information, can have predictive power. At minimal, technical analysis summarizes some time-series properties of stock prices, although their statistical properties are not well-studied.

Second, technical analysis is a study of market price adjustments. The equilibrium price adjusts to new information every day. The adjustment reflects shifts in demand and supply as a result of the reactions of all market participants to new information. However, this adjustment can be sluggish due to various reasons such as noise, market frictions, and investors' herding behavior. Noisy rational expectations models and feedback models, among others, attempt to provide theoretical justification for sluggish price adjustments. For example, see Working (1958), Brown and Jennings (1989), and Wang (1993) for noisy rational expectations model; De Long *et al.* (1990) and Shleifer and Summers (1990) for feedback models. The common implication of these models is that systematic adjustments in stock prices induce short-term serial correlations. That is, the current price need not fully reveal all available information. Therefore, technical signals may indicate the likely direction of price adjustment.

[3] For example, Bossaerts and Hillon (1990) provide strong evidence in favor of in-sample return predictability using an international stock market data set, although they find weak evidence for out-of-sample predictability. However, Cochrane (2008) argues that poor out-of-sample performance is not a test against predictability.

Third, showing predictive power of a trading rule does not itself rebut market efficiency[4]. Forecast errors can be large regardless of the quality of estimation. Therefore, the investor's portfolio need not generate any significant abnormal return after adjustment for risk and transaction costs. However, an investor may disadvantage himself if he completely ignores technical signals. For example, Hand, Yang, and Zhou (2013) show that some technical signals contain unique economic information that is not already contained in other information sources.

Fourth, investors may well recognize that technical signals need not be the most powerful class of predictive variables for stock returns[5], but other variables are typically not observed frequent enough for real-time trading. In particular, technical signals can be valuable to investors who engage in high frequency trading, for example, investment banks, pension funds, mutual funds, and other buy-side institutional traders.

There are two features about portfolio choice with a prediction model that are worth emphasizing. First, the true values of the model parameters are not necessarily relevant to an investor with a finite horizon. To be precise, consider the following linear prediction model for stock returns: $R_{t+1} = \beta_0 + \beta_1 X_t + \varepsilon_t$, where X_t is a zero-one technical signal[6]. Given a sample of T observations of (R_{t+1}, X_t), let $(\hat{\beta}_0^T, \hat{\beta}_1^T)$ denote some estimators of (β_0, β_1). The true values of the model parameters are often justified by the probability limit given by $(\beta_0, \beta_1) = \text{plim}_{T\to\infty}(\hat{\beta}_0^T, \hat{\beta}_1^T)$. However, the finite sample estimates $(\hat{\beta}_0^T, \hat{\beta}_1^T)$ are the parameter values directly relevant for constructing forecasts, even though their

[4] Fama (1970) points out that the notion of market efficiency does not necessarily imply successive price changes to be independent.

[5] There are many well-studied predictive variables for returns, for example, dividend yields, earnings-price ratio, term spreads, and expected inflation.

[6] For example, a "buy" signal is represented by $X_t = 1$ and $X_t = 0$ for a "sell" signal.

values need not be close to the hypothetical true values.

Second, a sophisticated technical investor would consider estimation risk. Indeed, due to random variation of sampling, even if the estimators have nice statistical properties, the true parameters are never known with certainty unless one has an infinite sample. Under the Bayesian approach, the investor integrates his expected utility over the unknown parameter space for portfolio optimization. His optimal portfolio strategy incorporates the degree of uncertainty about the unknown parameters, often measured by the covariances of the estimates $(\hat{\beta}_0^T, \hat{\beta}_1^T)$. Studies sharing this idea include Bawa, Brown, and Klein (1979), Kandel and Stambaugh (1996), Brennan (1998), Barberis (2000), Jacquier, Kane, and Marcus (2005), and Xia (2001).

2.3 The Model

2.3.1 The Basic Setting

Consider an investor with a long horizon who trades continuously in a two-asset economy in which a risk-free asset pays an instantaneous rate of interest r, and a risky stock represents the aggregate equity market.

We fix a finite horizon $[0, T]$. The cum-dividend stock price grows according to the following process

$$\frac{\mathrm{d}P_t}{P_t} = \mu_t \mathrm{d}t + \sigma \mathrm{d}B_t^*,$$

where the percentage volatility σ is known to the investor due to observable quadratic variations of P_t;[7] B_t^* is a Brownian motion defined on the probabil-

[7] In practice, we do not have true continuous-time data and so σ has to be estimated subject to some error. We nonetheless make the assumption that the investor in this model economy

ity space $(\Omega, \mathbb{P}^*, \mathcal{F}^*)$ with a standard filtration $\mathcal{F}^* \equiv \{\mathcal{F}_t^* : t \leq T\}$. However, the instantaneous percentage drift, $\mu_t \in \mathcal{F}_t^*$, is unknown to the investor.

Let S and L, with $0 < S < L < T$, be two lookback periods. We define a *moving average signal*, denoted by X_t, as follows,

$$X_t = \begin{cases} 1 & \text{if } D_t^{S,L} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$D_t^{S,L} \equiv \frac{1}{S} \int_{t-S}^{t} P_\tau \mathrm{d}\tau - \frac{1}{L} \int_{t-L}^{t} P_\tau \mathrm{d}\tau$$

is the difference of two moving averages of stock prices. When $X_t = 1$, we say a *buy* signal is generated, and a *sell* signal when $X_t = 0$. We assume the history $\{P_t : -L \leq t \leq 0\}$ is known. It is also useful to note that the dynamics of $D_t^{S,L}$ are given by

$$\mathrm{d}D_t^{S,L} = \left[\frac{1}{S}(P_t - P_{t-S}) - \frac{1}{L}(P_t - P_{t-L}) \right] \mathrm{d}t.$$

Let $\mathbb{P}$ denote the investor's subjective probability measure and $\mathcal{F}_t \equiv \{P_\tau : \tau \leq t\}$, with $\mathcal{F}_t \subset \mathcal{F}_t^*$, denote the investor's information set at time t. The investor conjectures a linear predictor $\beta_0 + \beta_1 X_t$ of μ_t, and thus under the reference model probability $\mathbb{P}$ the dynamics of stock price are described by the following *linear prediction model*

$$\frac{\mathrm{d}P_t}{P_t} = (\beta_0 + \beta_1 X_t)\mathrm{d}t + \sigma \mathrm{d}B_t,$$

where (β_0, β_1) are unknown model parameters to be estimated; B_t is a standard $\mathbb{P}$-Brownian motion adapted to the investor's information set $\mathcal{F}_t$. While the same

does have access to continuous-time data. This assumption is also used by Brennan (1998) and Xia (2001), among others.

linear prediction model is also used in Xia (2001), we note that there are two important differences: first, X_t is a zero-one variable, not a continuous variable; second it is a function of past stock prices only with no assumption on its law of motion required, and indeed the common assumption that the predictive variable follows the Ornstein-Uhlenbeck process is not applicable (see also Campbell and Viceira 1999; Kim and Omberg 1996; Wachter 2002). Here, we also relax the assumption that the intercept is known.

In this article, we assume the linear structure is appropriate without considering non-linear settings. For this reason, the usefulness or the predictive power of X_t is solely determined by whether the slope parameter β_1 is nonzero. Hence, in this sense, the uncertainty predictive power of X_t is translated to the uncertainty of the parameter β_1.

For any fixed point of time t in $[0, T)$, given an initial wealth W_t and the investment horizon T, the investor chooses a portfolio allocation ξ to maximize his expected utility of wealth,

$$\max_{\xi} \mathrm{E}[U(W_T)|\mathcal{F}_t],$$

given the wealth dynamics

$$\frac{\mathrm{d}W_\tau}{W_\tau} = r\mathrm{d}\tau + \xi(\beta_0 + \beta_1 X_\tau - r)\mathrm{d}\tau + \xi\sigma\mathrm{d}B_\tau.$$

We use the notation τ above because the letter t is already used to denote the fixed chosen point t. The modeling feature that the investor has a fixed investment horizon will allow us to compare the investment behaviors of shorter and longer-run investors by varying T.

Note that ξ is a constant to solve at each point of time t. We call the sequence of solutions of ξ a *rolling strategy* because the investor is essentially solving a buy-and-hold portfolio choice problem[8] continuously. Since the conditional distribution of W_T varies over time due to continuously updated information, so does the maximizer

$$\xi_t^* \equiv \arg\max_{\xi} \mathrm{E}[U(W_T)|\mathcal{F}_t].$$

We call ξ_t^* the *optimal allocation*. Our behavioural assumption that the investor follows this rolling strategy may be strong. A more realistic model would allow for a dynamic strategy to account for the possibility of learning the model parameters in the future, as in Brennan (1998) and Xia (2001). However, we believe that this rolling assumption can also give use useful implications for portfolio choice and investor welfare because it allows us to obtain a more tractable solution for allocation to stocks.

In this article, we assume the power-utility function

$$U(W_T) = \frac{W_T^{1-\gamma}}{1-\gamma},$$

where γ is the investor's risk-aversion parameter. We shall only consider investors with a risk-aversion parameter greater than the logarithmic case (*i.e.*, $\gamma > 1$).

2.3.2 The Investor's Optimization Problem

It is useful to rewrite the utility function as

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

[8] Previous studies considering the buy-and-hold strategy include Stambaugh (1999), Barberis (2000), and Avramov (2002).

and apply Itô's rule to obtain $\log W_T$,

$$\log W_T = \log W_t + [r + \xi(\beta_0 + \beta_1 \overline{\mathcal{X}}_t^T - r) - \xi^2 \tfrac{\sigma^2}{2}](T - t) + \xi\sigma(B_T - B_t),$$

where we define the variable

$$\overline{\mathcal{X}}_t^T \equiv \frac{1}{T - t} \int_t^T X_\tau \mathrm{d}\tau,$$

which can be interpreted as the average value of the moving average signals over the period $[t, T]$. Since $(\beta_0, \beta_1, \overline{\mathcal{X}}_t^T)$ are not adapted to the investor's information set $\mathcal{F}_t$, they are considered random variables at time t.

Let $\theta_t \equiv \beta_0 + \beta_1 \overline{\mathcal{X}}_t^T$ denote the *average percentage drift*[9] over the remaining investment horizon. By the law of total expectation, we have

$$\mathrm{E}[U(W_T)|\mathcal{F}_t] = \int_{\mathbb{R}} \mathrm{E}[U(W_T)|\mathcal{F}_t, \theta_t]\phi(\theta_t|\mathcal{F}_t)\mathrm{d}\theta_t, \tag{2.1}$$

where $\phi(\theta_t|\mathcal{F}_t)$ is the conditional density of θ_t and the integral is taken over the real line $\mathbb{R}$. Neglecting the potential correlation between θ_t and $B_T - B_t$,[10] by log-normality, the expectation inside the integral is

$$\mathrm{E}[U(W_T)|\mathcal{F}_t, \theta_t] = e^{r(1-\gamma)(T-t)} U(W_t) \exp\left\{(1-\gamma)\xi(\theta_t - r - \xi\tfrac{\gamma\sigma^2}{2})(T-t)\right\}.$$

Suppose that θ_t follows a conditional Gaussian distribution with mean $m_t \equiv$

[9] Note that we can write $\theta_t = \frac{1}{T-t}\int_t^T (\beta_0 + \beta_1 X_\tau)\mathrm{d}\tau$.

[10] In general, θ_t and $B_T - B_t$ are correlated. However, after calibrating our model with real data, our simulation study shows that their correlation is very little (less than 10^{-3} across time). In Appendix 2.A, we provide a discussion when their correlation is accounted for and show that ξ_t^* depends on the covariance between θ_t and $B_T - B_t$. We also analytically derive bounds on the covariance, which can useful to formulate a rough approximation for the covariance to reduce computational effort.

$\mathrm{E}[\theta_t|\mathcal{F}_t]$ and variance $v_t \equiv \mathrm{Var}[\theta_t|\mathcal{F}_t]$, then we can write equation (2.1) as

$$\mathrm{E}[U(W_T)|\mathcal{F}_t] \propto \exp\left\{(1-\gamma)\xi\{(m_t - r) - \tfrac{\xi}{2}[\gamma\sigma^2 + (\gamma-1)v_t(T-t)]\}(T-t)\right\}, \tag{2.2}$$

where the notation "$\propto$" means "is proportional to". Even if θ_t is nonGaussian, we can approximate it by a Gaussian process that matches the first and the second moments. We take (2.2) as the true expression and optimize it with respect to ξ to obtain the optimal allocation

$$\xi_t^* = \frac{m_t - r}{\gamma\sigma^2 + (\gamma-1)v_t(T-t)}. \tag{2.3}$$

2.3.3 The Investor's Inference Problem

Let us use the notation $\beta \equiv (\beta_0, \beta_1)$. By the law of total expectation, we can evaluate m_t as follows

$$\begin{aligned} m_t \equiv \mathrm{E}[\theta_t|\mathcal{F}_t] &= \mathrm{E}^{\beta}\{\mathrm{E}^{\overline{\mathcal{X}}}[\beta_0 + \beta_1\overline{\mathcal{X}}_t^T|\mathcal{F}_t, \beta]|\mathcal{F}_t\} \\ &= \mathrm{E}[\beta_0|\mathcal{F}_t] + \mathrm{E}[\beta_1\mathcal{E}_t^{\overline{\mathcal{X}}}(\beta)|\mathcal{F}_t], \end{aligned}$$

where we define the function $\mathcal{E}_t^{\overline{\mathcal{X}}} : \mathbb{R}^2 \to \mathbb{R}$ by

$$\mathcal{E}_t^{\overline{\mathcal{X}}}(\beta) \equiv \mathrm{E}[\overline{\mathcal{X}}_t^T|\mathcal{F}_t, \beta],$$

which is a function of β to be found. The operator $\mathrm{E}^{\beta}\{\cdot\}$ indicates that the expectation is taken over the distribution of β, and similarly, $\overline{\mathcal{X}}_t^T$ for $\mathrm{E}^{\overline{\mathcal{X}}}[\cdot]$. In the reference model $\mathbb{P}$, the random process $\{X_\tau : t < \tau \le T\}$ is determined by $\{P_\tau : t \le \tau \le T\}$, which depends on β, thus the expectation $\mathrm{E}[\overline{\mathcal{X}}_t^T|\mathcal{F}_t, \beta]$ also depends on β and cannot be factored out from the outer expectation $\mathrm{E}^{\beta}\{\cdot\}$. By contrast, if the law of motion of X_t were not determined by the prediction model itself, we would simply have $m_t = \mathrm{E}[\beta_0|\mathcal{F}_t] + \mathrm{E}[\beta_1|\mathcal{F}_t] \cdot \mathrm{E}[\overline{\mathcal{X}}_t^T|\mathcal{F}_t]$, which would be linear in the conditional expectations of β and we could obtain a closed-form solution for $\mathrm{E}[\overline{\mathcal{X}}_t^T|\mathcal{F}_t]$ if X_t followed the Ornstein-Uhlenbeck process. Thus, our question at hand involves a mathematical challenge not discussed in previous studies.

Similarly, by the law of total variance, we can evaluate v_t as follows,

$$\begin{aligned} v_t \equiv \mathrm{Var}[\theta_t|\mathcal{F}_t] &= \mathrm{E}^\beta\{\mathrm{Var}^{\overline{\mathcal{X}}}[\beta_0 + \beta_1\overline{\mathcal{X}}_t^T|\mathcal{F}_t, \beta]|\mathcal{F}_t\} \\ &\quad + \mathrm{Var}^\beta\{\mathrm{E}^{\overline{\mathcal{X}}}[\beta_0 + \beta_1\overline{\mathcal{X}}_t^T|\mathcal{F}_t, \beta]|\mathcal{F}_t\} \\ &= \mathrm{E}[\beta_1^2\mathcal{V}_t^{\overline{\mathcal{X}}}(\beta)|\mathcal{F}_t] + \mathrm{Var}[\beta_0 + \beta_1\mathcal{E}_t^{\overline{\mathcal{X}}}(\beta)|\mathcal{F}_t], \end{aligned}$$

where we define the function $\mathcal{V}_t^{\overline{\mathcal{X}}} : \mathbb{R}^2 \to \mathbb{R}^+$ by

$$\mathcal{V}_t^{\overline{\mathcal{X}}}(\beta) \equiv \mathrm{Var}[\overline{\mathcal{X}}_t^T|\mathcal{F}_t, \beta],$$

which is a function of β to be found. The operator $\mathrm{Var}^\beta\{\cdot\}$ indicates that the variance is taken over the distribution of β, and similarly, $\overline{\mathcal{X}}_t^T$ for $\mathrm{Var}^{\overline{\mathcal{X}}}[\cdot]$.

Let $m_t^\beta \equiv \mathrm{E}[\beta|\mathcal{F}_t]$ and $v_t^\beta \equiv \mathrm{V}[\beta|\mathcal{F}_t]$ denote the conditional mean and the variance-covariance matrix of the investor's estimate of β. We assume that the investor has a bivariate Gaussian prior distribution, with mean vector m_0^β and variance-covariance matrix v_0^β. Henceforth, we call (m_0^β, v_0^β) the hyperparameters. Following Liptser and Shiryaev (2001), the distribution of β conditional on $\mathcal{F}_t$ is also Gaussian with mean vector m_t^β and variance-covariance matrix v_t^β.

Proposition 2.1. *Given that the prior probability distribution of β is Gaussian with mean vector m_0^β and variance-covariance matrix v_0^β, the solutions to m_t^β and*

v_t^{β} *are*

$$m_t^{\beta} = \left[I + \frac{v_0^{\beta}}{\sigma^2}\int_0^t (\vec{X}_\tau \vec{X}_\tau^\top)\mathrm{d}\tau\right]^{-1}\left[m_0^{\beta} + \frac{v_0^{\beta}}{\sigma^2}\int_0^t \vec{X}_\tau \frac{\mathrm{d}P_\tau}{P_\tau}\right],$$

$$v_t^{\beta} = \left[I + \frac{v_0^{\beta}}{\sigma^2}\int_0^t (\vec{X}_\tau \vec{X}_\tau^\top)\mathrm{d}\tau\right]^{-1} v_0^{\beta},$$

where $\vec{X}_t \equiv (1, X_t)^\top$ *is a* 2×1 *vector;* I *is a* 2×2 *identity matrix.*

We use the following notation to denote the elements of m_t^{β} and v_t^{β},

$$m_t^{\beta} = \begin{bmatrix} m_t^{\beta_0} \\ m_t^{\beta_1} \end{bmatrix}, \quad v_t^{\beta} = \begin{bmatrix} v_t^{\beta_0} & v_t^{\beta_0,\beta_1} \\ v_t^{\beta_0,\beta_1} & v_t^{\beta_1} \end{bmatrix}.$$

Also, we use $\partial_i[f] \equiv \partial f/\partial \beta_i$ and $\partial_{i,j}[f] \equiv \partial^2 f/\partial\beta_i\partial\beta_j$, $i, j \in \{0, 1\}$, to denote the partial derivatives of a given function f.

Proposition 2.2. *Suppose that the functions* $\mathcal{E}_t^{\overline{\mathcal{X}}}$ *and* $\mathcal{V}_t^{\overline{\mathcal{X}}}$ *are continuously differentiable around* m_t^{β}*, then* m_t *can be expressed as the following Taylor series,*

$$\begin{aligned} m_t = {} & [m_t^{\beta_0} + m_t^{\beta_1}\mathcal{E}_t^{\overline{\mathcal{X}}}] + (\tfrac{1}{2}m_t^{\beta_1}\partial_{0,0}[\mathcal{E}_t^{\overline{\mathcal{X}}}])v_t^{\beta_0} + (\partial_1[\mathcal{E}_t^{\overline{\mathcal{X}}}] + \tfrac{1}{2}m_t^{\beta_1}\partial_{1,1}[\mathcal{E}_t^{\overline{\mathcal{X}}}])v_t^{\beta_1} \\ & + (\partial_0[\mathcal{E}_t^{\overline{\mathcal{X}}}] + m_t^{\beta_1}\partial_{0,1}[\mathcal{E}_t^{\overline{\mathcal{X}}}])v_t^{\beta_0,\beta_1} + H_t, \end{aligned} \tag{2.4}$$

and v_t *can be expressed as the following Taylor series,*

$$\begin{aligned} v_t = {} & [(m_t^{\beta_1})^2\mathcal{V}_t^{\overline{\mathcal{X}}}] + \{\tfrac{1}{2}(m_t^{\beta_1})^2\partial_{0,0}[\mathcal{V}_t^{\overline{\mathcal{X}}}] + (1 + m_t^{\beta_1}\partial_0[\mathcal{E}_t^{\overline{\mathcal{X}}}])^2\}v_t^{\beta_0} \\ & + \{\mathcal{V}_t^{\overline{\mathcal{X}}} + 2m_t^{\beta_1}\partial_1[\mathcal{V}_t^{\overline{\mathcal{X}}}] + \tfrac{1}{2}(m_t^{\beta_1})^2\partial_{1,1}[\mathcal{V}_t^{\overline{\mathcal{X}}}] + (\mathcal{E}_t^{\overline{\mathcal{X}}} + m_t^{\beta_1}\partial_1[\mathcal{E}_t^{\overline{\mathcal{X}}}])^2\}v_t^{\beta_1} \\ & + \{2m_t^{\beta_1}\partial_0[\mathcal{V}_t^{\overline{\mathcal{X}}}] + (m_t^{\beta_1})^2\partial_{0,1}[\mathcal{V}_t^{\overline{\mathcal{X}}}] \\ & \quad + 2(1 + m_t^{\beta_1}\partial_0[\mathcal{E}_t^{\overline{\mathcal{X}}}])(\mathcal{E}_t^{\overline{\mathcal{X}}} + m_t^{\beta_1}\partial_1[\mathcal{E}_t^{\overline{\mathcal{X}}}])\}v_t^{\beta_0,\beta_1} + O_t, \end{aligned} \tag{2.5}$$

where $\mathcal{E}_t^{\overline{\mathcal{X}}}$, $\mathcal{V}_t^{\overline{\mathcal{X}}}$*, and their partial derivatives are evaluated at* m_t^{β}*;* H_t *and* O_t *are* *remainder terms.*

We give the proofs of these propositions in Appendix 2.B.

To our knowledge, closed-form solutions for $\mathcal{E}_t^{\overline{\mathcal{X}}}$ and $\mathcal{V}_t^{\overline{\mathcal{X}}}$ are not feasible. Therefore, we develop estimators and use Monte Carlo simulation to estimate these functions. Given any arbitrary integer n, we define the quantities $\Delta t \equiv T/n$, $\ell \equiv \lfloor nL/T \rfloor$, and $s \equiv \lfloor nS/T \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function. For a sufficiently large n, we have $L \simeq \ell \Delta t$ and $S \simeq s\Delta t$, where the notation "$\simeq$" means "is approximately equal to or equal to". We partition the time interval $[-\ell\Delta t, T]$ into $\ell + n$ equidistant subintervals,

$$-L \simeq t_{-\ell} < \cdots - S \simeq t_{-s} < \cdots 0 = t_0 < \cdots t_n = T,$$

and approximate any t in $[-L, T]$ by t_j if $t_{j-1} < t \leq t_j$, $j = -\ell, \ldots, n$. Next, given any value of β, we consider the following Euler difference scheme:

$$\hat{P}_{t_{j+1}} = \hat{P}_{t_j} + \hat{P}_{t_j}[(\beta_0 + \beta_1 \hat{X}_{t_j})\Delta t + \sigma\sqrt{\Delta t}\,\epsilon_{t_j}], \tag{2.6}$$

$$\hat{D}^{S,L}_{t_{j+1}} = \hat{D}^{S,L}_{t_j} + \left[\frac{1}{s}(\hat{P}_{t_j} - \hat{P}_{t_{j-s}}) - \frac{1}{\ell}(\hat{P}_{t_j} - \hat{P}_{t_{j-\ell}})\right], \tag{2.7}$$

$$\hat{X}_{t_{j+1}} = \begin{cases} 1 & \text{if } \hat{D}^{S,L}_{t_{j+1}} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.8}$$

with $\hat{P}_{t_j} = P_{t-j\Delta t}$ if $j \leq i$, $i = 0, \ldots, n-1$; ϵ_{t_j} are independent standard Gaussian random variables.

To simulate a trajectory of $\{\hat{X}_{t_j} : j = i, \ldots, n-1\}$, $i = 0, \ldots, n-1$, we start from the initial values $(\hat{P}_{t_i} = P_t; \hat{X}_{t_i} = X_t)$ and proceed recursively according to the Euler difference scheme (6)-(8). We then calculate the sum

$$\hat{\overline{\mathcal{X}}}_t^T \equiv \frac{1}{T-t} \sum_{j=i}^{n-1} \hat{X}_{t_j} \Delta t$$

to approximate the stochastic integral $\overline{\mathcal{X}}_t^T$. Let $\hat{\overline{\mathcal{X}}}_{t,k}^T$ denote the sum calculated using the k-th trajectory, $k = 1, \ldots, K$. To estimate the moment $\mathcal{E}_t^{\overline{\mathcal{X}}}(\beta)$, we use the estimator

$$\hat{\mathcal{E}}_t^{\overline{\mathcal{X}}}(\beta; \vec{\epsilon}_t) \equiv \frac{1}{K} \sum_{k=1}^{K} \hat{\overline{\mathcal{X}}}_{t,k}^T,$$

and similarly,

$$\hat{\mathcal{V}}_t^{\overline{\mathcal{X}}}(\beta; \vec{\epsilon}_t) \equiv \frac{1}{K} \sum_{k=1}^{K} (\hat{\overline{\mathcal{X}}}_{t,k}^T)^2 - \left[\hat{\mathcal{E}}_t^{\overline{\mathcal{X}}}(\beta; \vec{\epsilon}_t)\right]^2$$

for the moment $\mathcal{V}_t^{\overline{\mathcal{X}}}(\beta)$, where $\vec{\epsilon}_t = (\{\epsilon_{t_j,1}\}_{j=i}^{n-1}, \ldots, \{\epsilon_{t_j,K}\}_{j=i}^{n-1})$ is an $(n-i) \times K$ matrix of independent standard Gaussian random variables. We use the argument $(\,\cdot\,; \vec{\epsilon}_t)$ to emphasize that the values of $(\hat{\mathcal{E}}_t^{\overline{\mathcal{X}}}, \hat{\mathcal{V}}_t^{\overline{\mathcal{X}}})$ depend on $\vec{\epsilon}_t$. By fixing $\vec{\epsilon}_t$ for each t, the estimators $(\hat{\mathcal{E}}_t^{\overline{\mathcal{X}}}, \hat{\mathcal{V}}_t^{\overline{\mathcal{X}}})$ are smooth in β. Thus, we can evaluate the partial derivatives in the Taylor series (2.4)-(2.5) using finite difference approximation. We present the required formulas in Appendix 2.C.

2.4 Data, Model Calibration, and Some Empirical Facts

We measure the horizon T in years and consider daily intervals (with step size $\Delta = \frac{1}{252}$) small enough for good discrete approximations to the continuous-time processes in the model. We use the inflation adjusted CRSP index (including distributions) on a value-weighted basket of stocks listed in NYSE, AMEX, and NASDAQ as an example of the risky stock. We obtain the nominal daily index data for the period 2nd January 1980 to 31st December 2014 from the CRSP VWRETD file under the folder named "Index/Stock File Indexes". Then, we make a month-by-month inflation adjustment using the consumer price index (CPI), with January 1980 being the base month. The CPI data are obtained from the CRSP inflation file under the folder named "US Treasury and Inflation Indexes". The standard deviation of the daily inflation adjusted index returns is 1.083%. Dividing this figure by $\sqrt{\Delta t}$ gives $\sigma = 0.172$. We also obtain the nominal average interest rate data from the CRSP 30-day Treasury bill file. The monthly average interest rate is 0.382%, and 0.267% for the monthly average of the rate of change in CPI. Thus, the implied daily real interest rate is 0.004%. Diving this figure by Δt gives $r = 0.010$. For the Euler difference scheme, we use $K = 2000$ trajectories, although we find that even $K = 500$ would give similar results.

As an example, we consider 1-day and 100-day moving averages. In terms of the model notation, we have $s = 1$ and $\ell = 100$. To obtain reasonable parameters of the hyperparameters, namely (m_0^β, v_0^β), we use the inflation adjusted index data for the subsample period 2nd January 1980 to 22nd December 1994 (leaving 20×252 days for the out-of-sample period) to run the regression:

$$\frac{P_{t_{j+1}} - P_{t_j}}{P_{t_j}} = (\beta_0 + \beta_1 X_{t_j})\Delta t + \varepsilon_{t_j},$$

where $X_{t_j} = 1$ if the 1-day moving average is above the 100-day moving average at time t_j, and $X_{t_j} = 0$ otherwise. We then take the least-squares estimates as the hyperparameters,

$$m_0^\beta = \begin{bmatrix} 0.058 \\ 0.061 \end{bmatrix}, \quad v_0^\beta = \begin{bmatrix} 0.004 & -0.004 \\ -0.004 & 0.006 \end{bmatrix}.$$

Figure 2.1 plots the time series of the intercept and the slope estimates $(m_t^{\beta_0}, m_t^{\beta_1})$, when the investor starts investing on 23rd December 1994 with the hyperparameters $(m_0^{\beta}, v_0^{\beta})$. These time series end on the last trading day of year 2014. Observe that the slope estimates roughly follow a downward trend but clearly stay above zero. Note also that investors do expect to earn higher returns on buy-signal days than on sell-signal days (*i.e.*, $m_t^{\beta_0} + m_t^{\beta_1} > m_t^{\beta_0}$). Thus, it is not without ground that investors may believe that the technical signals contain useful information to forecast returns (even if the true but unknown value of β_1 is actually zero).

One may wonder whether the slope estimates remain nonzero because of the prior. While we cannot explain with full generality, we would like to make a conjecture. Suppose instead the investor has a short memory, for example, he uses the rolling estimation strategy instead of updating the prior from time zero. This is so-called short memory because any data collected before the rolling window are discarded. It appears, however, that the slope estimates would still remain nonzero for a reasonably long period. The rationale is as follows. Suppose we begin with the (strong) prior that the slope is zero, the data will still push the estimate up to a level similar to the OLS priors (because the sample period does clearly suggest a nonzero slope). Then we observe that although the Bayesian slope is descending, the process is rather slow. Indeed, there are a couple of clear bounces during the last twenty years. Hence, while the rolling estimation will speed up the descent of the slope estimate, with the presence of several bounces, it is still not easy to hit the zero level within a short period.

2.5 Analysis and Results

2.5.1 Mean and Variance Decompositions

We can express the expectation and the variance of the average percentage drift, (m_t, v_t), as follows

$$m_t = \tilde{m}_t + \pi_t^0 v_t^{\beta_0} + \pi_t^1 v_t^{\beta_1} + \pi_t^{01} v_t^{\beta_0,\beta_1} + H_t, \tag{2.9}$$

$$v_t = \tilde{v}_t + \lambda_t^0 v_t^{\beta_0} + \lambda_t^1 v_t^{\beta_1} + \lambda_t^{01} v_t^{\beta_0,\beta_1} + O_t. \tag{2.10}$$

The remainders H_t and O_t correspond to the ones in Proposition 2.2, and the definitions of the coefficients follow accordingly: $\tilde{m}_t \equiv m_t^{\beta_0} + m_t^{\beta_1} \mathcal{E}_t^{\overline{\mathcal{X}}}$, $\tilde{v}_t \equiv (m_t^{\beta_1})^2 \mathcal{V}_t^{\overline{\mathcal{X}}}$, and so on. We call $(\tilde{m}_t, \tilde{v}_t)$ the *basic component* of (m_t, v_t). We also call $(\tilde{m}_t, \tilde{v}_t)$ the *estimation-risk ignorant estimates* because they disregard any estimation risk adjustments, *i.e.*, the terms associated with $(v_t^{\beta_0}, v_t^{\beta_1}, v_t^{\beta_0,\beta_1})$ with each representing a source of estimation risk adjustment: the uncertainty in (β_0, β_1) individually and their joint-uncertainty since they have to be jointly estimated. For simplicity, we approximate (m_t, v_t) by summing only the first four terms in equations (2.9)-(2.10), *i.e.*, ignoring the remainders, although one could improve the approximation by including higher-order moments and derivatives.

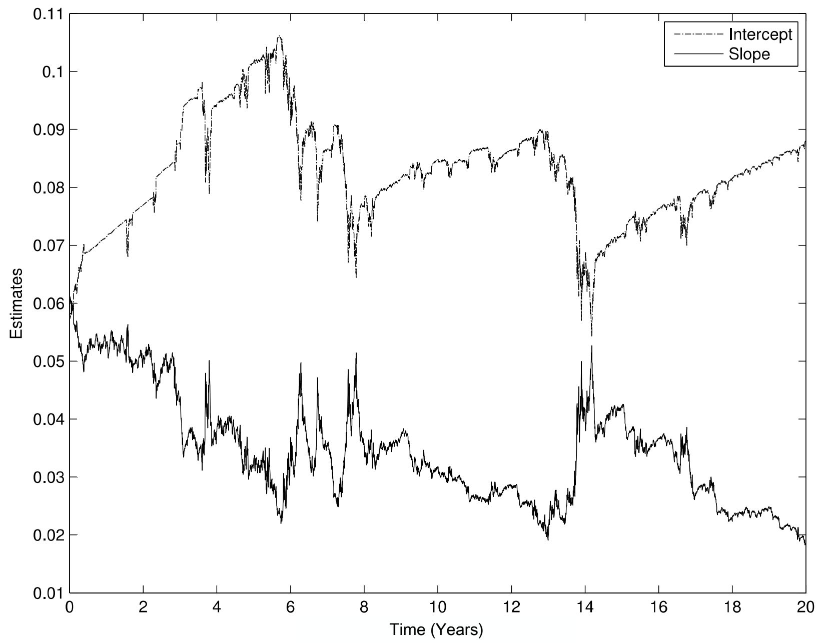

Figure 2.1. Estimates of intercept and slope parameters of the linear prediction model. This figure plots the time series of the intercept and the slope estimates, $(m_t^{\beta_0}, m_t^{\beta_1})$, when the investor starts investing on 23rd December 1994 until 31st December 2014 (20×252 trading days).

We divide equation (2.9) by m_t to obtain

$$\begin{aligned}\frac{m_t}{m_t} &= \frac{\tilde{m}_t}{m_t} + \frac{\pi_t^0 v_t^{\beta_0}}{m_t} + \frac{\pi_t^1 v_t^{\beta_1}}{m_t} + \frac{\pi_t^{01} v_t^{\beta_0,\beta_1}}{m_t} \\ 1 &= \tilde{M}_t + M_t^0 + M_t^1 + M_t^{01},\end{aligned}$$

where $\tilde{M}_t$ is the *percentage composition* of the basic component $\tilde{m}_t$; M_t^0, M_t^1, and M_t^{01}, for the estimation-risk adjustment components. Similarly, we divide equation (2.10) by v_t to obtain

$$\begin{aligned}\frac{v_t}{v_t} &= \frac{\tilde{v}_t}{v_t} + \frac{\lambda_t^0 v_t^{\beta_0}}{v_t} + \frac{\lambda_t^1 v_t^{\beta_1}}{v_t} + \frac{\lambda_t^{01} v_t^{\beta_0,\beta_1}}{v_t} \\ 1 &= \tilde{V}_t + V_t^0 + V_t^1 + V_t^{01},\end{aligned}$$

where $\tilde{V}_t$ is the percentage composition of the basic component $\tilde{v}_t$; V_t^0, V_t^1, and V_t^{01}, for the estimation-risk adjustment components.

We consider four investment horizons, $T = \{5, 10, 15, 20\}$. For each horizon, we perform the mean and variance decompositions at five points of time, $t = \delta T$, $\delta = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{251}{252}\}$. Panel A of Table 2.1 displays the results of the mean decomposition. We observe that the basic component accounts for roughly 100% of m_t most of the time while the estimation risk adjustment components are relatively small. Our results show that ignoring estimation risk does not materially bias the expectation of the average percentage drift θ_t. Panel B of Table 2.1 displays the results of the variance decomposition. We observe that the size of the estimation-risk ignorant component is materially smaller than that of the estimation-risk adjustment components. Notably, the estimation-risk ignorant component becomes essentially zero towards the end of the horizon. Our results show that ignoring estimation risk materially underestimates the variance of the average percentage drift θ_t.

Note that both (m_t, v_t) and $(\tilde{m}, \tilde{v}_t)$ depend on the technical trading rule and the stock price data. Therefore, their relative sizes can only be determined empirically. Without actual data, we would have an inconclusive theoretical discussion.

Table 2.1. Mean and Variance Decompositions of Average Percentage Drift

This table presents the results of the mean and variance decompositions of the average percentage drift θ_t. The expectation of θ_t is denoted by m_t; $\tilde{M}_t$ denotes the percentage composition of the basic component $\tilde{m}_t$; (M_t^0, M_t^1, M_t^{01}) denote the percentage compositions of the estimation risk adjustment components associated with $(v_t^{\beta_0}, v_t^{\beta_1}, v_t^{\beta_0,\beta_1})$. The variance of θ_t is denoted by v_t; $\tilde{V}_t$ denotes the percentage composition of the basic component $\tilde{v}_t$; (V_t^0, V_t^1, V_t^{01}) denote the percentage compositions of the estimation risk adjustment components associated with $(v_t^{\beta_0}, v_t^{\beta_1}, v_t^{\beta_0,\beta_1})$. The results show that ignoring estimate risk does not materially bias the expectation of θ_t but materially underestimates its variance. The figures are measured in percent. The notation "$*$" denotes values less than 5×10^{-5} percent.

	Panel A: Mean Decomposition						Panel B: Variance Decomposition				
	m_t	$\tilde{M}_t$	M_t^0	M_t^1	M_t^1	M_t^{01}	v_t	$\tilde{V}_t$	V_t^0	V_t^1	V_t^{01}
$t=0$											
T	5	9.246	100.996	−0.028	4.523	−5.492	0.170	2.394	272.706	141.971	−317.070
	10	9.218	101.522	−0.247	4.430	−5.705	0.164	1.238	283.437	147.039	−331.714
	15	9.239	101.352	−0.206	4.523	−5.668	0.162	0.833	286.983	148.586	−336.401
	20	9.259	101.173	−0.164	4.590	−5.599	0.161	0.630	289.006	149.615	−339.251
$t=\frac{1}{4}T$											
T	5	10.483	100.703	−0.027	3.867	−4.543	0.158	2.502	273.141	158.191	−333.834
	10	11.248	100.650	−0.101	3.450	−3.999	0.141	1.174	293.917	171.028	−366.119
	15	11.148	100.650	−0.027	3.303	−3.926	0.133	0.627	299.737	160.516	−360.881
	20	12.099	100.529	−0.061	2.936	−3.405	0.120	0.285	309.165	179.205	−388.656
$t=\frac{1}{2}T$											
T	5	11.377	100.508	−0.076	3.315	−3.747	0.151	3.132	273.353	175.882	−352.368
	10	12.134	100.487	−0.076	2.892	−3.303	0.124	0.818	299.676	181.614	−382.108
	15	9.997	100.822	−0.099	2.918	−3.642	0.115	0.771	276.053	144.729	−321.552
	20	10.358	100.418	−0.027	2.748	−3.138	0.102	0.503	280.635	161.321	−342.459
$t=\frac{3}{4}T$											
T	5	10.814	100.470	0.231	2.780	−3.480	0.167	3.766	234.774	103.979	−242.518
	10	9.845	101.110	−0.140	2.587	−3.557	0.126	2.045	250.804	121.513	−274.361
	15	10.234	100.277	0.049	2.704	−3.031	0.101	0.908	272.862	158.706	−332.477
	20	9.742	100.204	0.021	2.579	−2.804	0.094	1.999	265.410	154.560	−321.969
$t=\frac{251}{252}T$											
T	5	13.348	100.000	$*$	$*$	$*$	0.157	$*$	219.977	319.977	−439.954
	10	11.645	100.000	$*$	$*$	$*$	0.139	$*$	191.320	291.320	−382.640
	15	11.412	100.000	$*$	$*$	$*$	0.122	$*$	185.458	285.458	−370.916
	20	10.649	100.000	$*$	$*$	$*$	0.105	$*$	198.146	298.146	−396.292

Table 2.2. Horizon Effect and Hedging Demand against Estimation Risk

This table presents the optimal and suboptimal allocations to stocks and their difference as the hedging demand against estimation risk. The optimal allocation is defined by $\xi_t^* \equiv (m_t - r)/[\gamma\sigma^2 + (\gamma - 1)v_t(T - t)]$ and the suboptimal allocation is defined by $\tilde{\xi}_t \equiv (\tilde{m}_t - r)/[\gamma\sigma^2 + (\gamma - 1)\tilde{v}_t(T - t)]$, where T and γ are the investor's investment horizon and risk-aversion parameter. The hedging demand is defined by $\Delta_t \equiv \xi_t^* - \tilde{\xi}_t$. We consider four investment horizons, $T = \{5, 10, 15, 20\}$, and present $(\xi_t^*, \tilde{\xi}_t, \Delta_t)$ at five points of time, $t = \delta T$, $\delta = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{251}{252}\}$. The results show that, at time $t = 0$, the optimal allocation monotonically decreases with horizon, while the suboptimal allocation is not sensitive to the horizon. As the time to horizon decreases, the hedging demand decreases and eventually becomes close to zero. The figures are measured in percent.

	Optimal Strategy				Suboptimal Strategy				Difference			
T	5	10	15	20	5	10	15	20	5	10	15	20
$t = 0$												
γ 5	45.606	38.736	33.829	30.049	56.392	56.531	56.565	56.589	−10.787	−17.796	−22.736	−26.540
7	32.147	27.075	23.499	20.774	40.264	40.364	40.388	40.405	−8.118	−13.289	−16.890	−19.631
9	24.821	20.810	18.001	15.874	31.310	31.387	31.406	31.419	−6.489	−10.577	−13.405	−15.545
$t = \frac{1}{4}T$												
γ 5	55.548	54.121	49.116	50.645	64.691	69.889	69.263	75.714	−9.143	−15.768	−20.146	−25.069
7	39.290	38.052	34.377	35.346	46.195	49.909	49.465	54.076	−6.905	−11.857	−15.088	−18.730
9	30.394	29.341	26.442	27.146	35.924	38.813	38.469	42.057	−5.530	−9.472	−12.027	−14.911
$t = \frac{1}{2}T$												
γ 5	63.969	64.753	49.594	49.833	70.665	75.917	61.607	63.805	−6.696	−11.164	−12.013	−13.972
7	45.392	45.782	34.952	35.053	50.464	54.221	43.999	45.571	−5.072	−8.439	−9.047	−10.518
9	35.176	35.408	26.985	27.035	39.244	42.170	34.219	35.442	−4.068	−6.761	−7.234	−8.407
$t = \frac{3}{4}T$												
γ 5	63.103	55.407	56.918	52.733	66.888	60.765	62.899	59.421	−3.784	−5.359	−5.981	−6.688
7	44.901	39.355	40.387	37.366	47.770	43.399	44.925	42.436	−2.868	−4.043	−4.538	−5.070
9	34.849	30.515	31.297	28.934	37.151	33.752	34.940	33.002	−2.302	−3.237	−3.643	−4.069
$t = \frac{251}{252}T$												
γ 5	83.809	72.292	70.716	65.557	83.837	72.314	70.734	65.572	−0.028	−0.022	−0.018	−0.015
7	59.862	51.636	50.510	46.826	59.883	51.653	50.524	46.837	−0.022	−0.016	−0.014	−0.011
9	46.559	40.161	39.285	36.420	46.576	40.174	39.297	36.429	−0.017	−0.013	−0.011	−0.009

2.5.2 Horizon Effect and Hedging Demand

Let us define

$$\tilde{\xi}_t \equiv \frac{\tilde{m}_t - r}{\gamma\sigma^2 + (\gamma - 1)\tilde{v}_t(T - t)}. \tag{2.11}$$

We call $\tilde{\xi}_t$ the *suboptimal allocation* in the sense that it is based on the estimation-risk ignorant estimates $(\tilde{m}_t, \tilde{v}_t)$, not the optimal estimates (m_t, v_t). Equivalently, we obtain $\tilde{\xi}_t$ by setting $(v_t^{\beta_0},\ v_t^{\beta_1},\ v_t^{\beta_0,\beta_1})$ to zero in ξ_t^* given by (2.3). Both the optimal and the suboptimal allocations depend on the time to horizon, namely $T - t$: explicitly through the denominators; and implicitly through the expectation and the variance of the average percentage drift θ_t, namely $(m_t, \tilde{m}_t, v_t, \tilde{v}_t)$. The influence of the time to horizon on the portfolio allocation is commonly called the *horizon effect*. Also, we call the difference of the two allocations, denoted by $\Delta_t \equiv \xi_t^* - \tilde{\xi}_t$, the investor's *hedging demand*. We can interpret this quantity as an allocation to hedge against estimation risk.

We consider three values of the risk-aversion parameter, $\gamma = \{5, 7, 9\}$. For each investment horizon, $T = \{5, 10, 15, 20\}$, we calculate the optimal allocation, the suboptimal allocation, and the hedging demand at five points of time, $t = \delta T$, $\delta = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{251}{252}\}$. Table 2.2 displays our results. For comparison, we also compute the optimal constant allocation to stocks, namely $(\overline{\theta} - r)/(\gamma\sigma^2) = (56.602\%, 40.430\%, 31.446\%)$, for $\gamma = \{5, 7, 9\}$, where $\overline{\theta} = 0.093$ is the (full) sample average percentage drift. Let us consider time $t = 0$. Observe first that, for all values of γ, the optimal allocation monotonically decreases as the horizon increases while the suboptimal allocation is not sensitive to the horizon. This is because the empirical values of $\tilde{v}_t$ are sufficiently small that the adjustment term $\tilde{v}_t(T - t)$ makes a negligible contribution to $\tilde{\xi}_t$. Second, the optimal allocation decreases more sharply for higher value of γ as the horizon increases. Overall, the hedging demand is increasing in horizon but decreasing in risk-aversion parameter.

Next, we look at other points of time along the investment horizon. Our results show that the horizon effect remains clear. At time $t = \frac{3}{4}T$, with only a quarter of the horizon remaining, the values of the hedging demand are still above 2%. However, the hedging demand gradually declines to zero. The reason is that as the estimates become more precise and the time to horizon decreases, the compounding effect of estimation risk eventually vanishes.

The notion that investors facing estimation risk may allocate less to stocks for longer horizon is not new. For example, in the context of a constant-plus-noise model for returns, Brennan (1998) finds that the dynamic allocation declines monotonically as the horizon is reduced when the investor is more risk tolerant than a logarithmic investor; in the context of a predictive VAR model for returns, Barberis (2000) finds that the allocation can decrease with horizon when the investor adapts a buy-and-hold strategy. Stambaugh (1999) and Xia (2001) also show that the allocation can decline eventually when the horizon becomes sufficiently long, although the allocation can first increase with horizon. Our results are consistent with these findings.

2.5.3 Market Timing Effect

When stock returns are deemed to be predictable, the portfolio allocation can depend on the current value of the predictive variable. This is called the *market timing effect*. For our model, when the slope estimate is positive, then a buy signal implies a higher expected stock returns and thus attracts the investor to increase his allocation to stocks. To assess market timing effect, we regress allocation to stocks on the moving average signal and the time to horizon for each trading day from 23rd December 1994 to 20th December 1999 (5×252 days):

$$\xi_t = \alpha_0 + \alpha_1 X_t + \alpha_2 (T - t) + \varepsilon_t,$$

where α_1 measures the (average) difference in allocation to stocks between buy and sell-signal days (after controlling for a time trend). We note that this regression approach to measure market timing is proposed because otherwise ξ_t is a complicated function of X_t. Henceforth, we interpret α_1 as a measure of the average market timing effect. We choose this subsample period because it is the widest time interval in which our portfolios are subject to the same market fluctuations. During this period, the ratio of buy and sell signals is 960:300 (or 3.2:1), and the longest runs of buy signals last for 126, 139, and 143 days.

We again compare the optimal and suboptimal strategies for four investment horizons, $T = \{5, 10, 15, 20\}$, and three values of risk-aversion parameter, $\gamma = \{5, 7, 9\}$. We report the least-squares estimates of α_1 in Table 2.3. First, observe that α_1 is positive and it has a range of 0.785% to 4.825%. Since α_1 is far from 100%, this confirms our intuition that the all-or-nothing strategy is too aggressive to adapt. Second, we observe two patterns of α_1: it decreases as the investment horizon increases; and it also decreases as the risk-aversion parameter increases. This is consistent with our intuition that when estimation risk is account for, longer-horizon and more risk-averse investors are more conservative in market timing. Third, we note that α_1 associated with the optimal strategy decreases more notably with horizon relative to that of the suboptimal strategy. Figure 2.2 highlights the differences in α_1. For example, fixing $\gamma = 5$, we see that the difference in α_1 is only -0.062% for $T = 5$ but as T increases to 20, the difference becomes -0.973%.

Although our model differs from Zhu and Zhou (2009) in how the investor infers

Table 2.3. Average Market Timing Effect

We regress allocation to stocks on the moving average signal and time to horizon for each trading day from 23rd December 1994 to 20th December 1999 (5×252 days): $\xi_t = \alpha_0 + \alpha_1 X_t + \alpha_2(T - t) + \varepsilon_t$, where $X_t = 1$ if the 1-day moving average is above the 100-day moving average at time t, and $X_t = 0$ otherwise; α_1 measures the (average) difference in allocation to stocks between buy and sell-signal days (after isolating time trend). We interpret α_1 as a measure of average market timing effect. The least-squares estimates of α_1 are reported as follows. The results show that, on average, the investor allocates an additional proportion of his wealth to the stock when a buy signal is observed. The figures are measured in percent.

		Optimal Strategy				Suboptimal Strategy				Difference			
	T	5	10	15	20	5	10	15	20	5	10	15	20
γ	5	4.763	2.152	1.734	1.484	4.825	2.678	2.518	2.456	−0.062	−0.526	−0.784	−0.972
	7	3.381	1.509	1.207	1.027	3.445	1.912	1.798	1.754	−0.064	−0.404	−0.591	−0.727
	9	2.620	1.161	0.926	0.785	2.679	1.487	1.398	1.364	−0.059	−0.326	−0.473	−0.579

information from the moving average signals, we also find that it is optimal for the investor to increase his allocation to stocks when he observes a buy signal. Our model, however, allows the investor to continuously assess the predictive power of the signals and adjusts his allocations to stocks accordingly. In particular, while the slope estimate remains positive in our sample period, it could be negative. In that case, the investor would actually decrease his allocation to stocks even though a "buy" signal is observed. This is possible because if the current price is well above the moving averages, the chance of a price reversal is also high. Therefore, the sign of the slope estimate depends on whether the data suggest that the buy signals are more often associated with upward trends or price reversals on average.

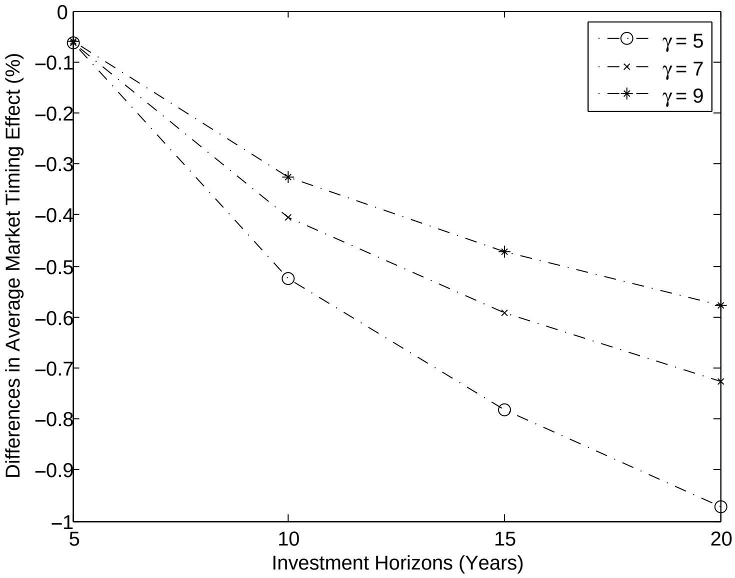

Figure 2.2. Differences in average market timing effect between the optimal and the suboptimal allocation strategies. This figure plots the difference in the average market timing effect between the optimal and the suboptimal allocations (reported in Table 2.3) as a function of the investment horizon T and the risk-aversion parameter γ. This figure highlights that the average market timing effect is stronger for the suboptimal strategy, and the difference becomes more notable as the horizon increases and the risk-aversion parameter decreases. The figures are measured in percent.

2.5.4 Welfare Costs of Ignoring Estimation Risk

Given an allocation ξ_t to the stock at time t, we define a function J by

$$\begin{aligned}&J(t, W_t, m_t, v_t; \xi_t)\\ &\qquad \equiv U(W_t) \exp\left\{(1-\gamma)\xi_t\{(m_t - r) - \tfrac{\xi_t}{2}[\gamma\sigma^2 + (\gamma - 1)v_t(T-t)]\}(T-t)\right\},\end{aligned}$$

which is the investor's maximized expected utility at time t. Next, we define the *welfare cost* (a measure of opportunity cost) of using the suboptimal allocation $\tilde{\xi}_t$

by the quantity ∇_t that satisfies the equality

$$J(t,1,m_t,v_t;\xi_t^*) = J(t,1+\nabla_t,m_t,v_t;\tilde{\xi}_t).$$

The welfare cost ∇_t represents the percentage wealth compensation required to leave the investor, who has \$1 today to invest up to the horizon, indifferent between choosing ξ_t^* and $\tilde{\xi}_t$. Similar measures are used in the literature (see, *e.g.*, Campbell and Viceira, 1999; Xia 2001; Han, Yang, and Zhou, 2013). By construction, J is maximized at $\xi_t = \xi_t^*$, and so the welfare cost is always nonnegative (*i.e.*, $\nabla_t \geq 0$). We are interested to compute ∇_t to assess how costly is it for an investor to ignore estimation risk.

We again consider four investment horizons, $T = \{5, 10, 15, 20\}$, and three values of risk-aversion parameter, $\gamma = \{5, 7, 9\}$. For each horizon, we compute the welfare cost at five points of time, $t = \delta T$, $\delta = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{251}{252}\}$. Table 2.4 displays the results. We see two patterns of welfare cost. First, it increases substantially as the horizon increases. For the shortest horizon considered (*i.e.*, $T = 5$), at time $t = 0$, its range is 0.352% to 0.530%. By contrast, for the longer horizons, $T = \{15, 20\}$, the range is 6.394% to 21.474%, and the welfare cost remains above 3% for at least one quarter of the horizon. Our results imply that it is costly for longer-horizon investors to ignore estimation risk. This is because the suboptimal strategy results in more severe overinvestment for longer horizons. By contrast, estimation risk is less of a concern for shorter-horizon investors. To further signify the magnitude of welfare cost, we fix the time at $t = 0$ and plot it as a function of horizon and risk-aversion parameter in Figure 2.3. The results show that the welfare cost increases with horizon at an increasing rate. The second pattern we observe is that the welfare cost increases as the risk-aversion parameter decreases and the increase is stronger for longer horizons. For example, for

Table 2.4. Welfare Cost for Ignoring Estimation Risk

This table presents the welfare cost which measures the percentage wealth compensation required to leave the investor, who has \$1 today to invest up to the horizon, indifferent between choosing the optional and the suboptimal allocation strategies. We consider four investment horizons, $T = \{5, 10, 15, 20\}$, and present the results at five points of time, $t = \delta T$, $\delta = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{251}{252}\}$. At time $t = 0$, the welfare cost is strongly increasing in horizon. For the longer horizons, $T = \{15, 20\}$, the welfare cost remains above 3% for at least one quarter of the horizon. The results show that it can be costly for a long-horizon investor to ignore estimation risk. The figures are measured in percent. The notation "$*$" denotes values less than 5×10^{-5} percent.

	T	5	10	15	20
$t = 0$					
γ	5	0.530	3.434	9.954	21.474
	7	0.426	2.730	7.829	16.644
	9	0.352	2.245	6.394	13.467
$t = \frac{1}{4}T$					
γ	5	0.269	1.789	4.851	10.927
	7	0.217	1.436	3.869	8.648
	9	0.180	1.187	3.186	7.084
$t = \frac{1}{2}T$					
γ	5	0.091	0.540	0.992	1.859
	7	0.074	0.436	0.797	1.495
	9	0.061	0.362	0.660	1.237
$t = \frac{3}{4}T$					
γ	5	0.014	0.058	0.109	0.187
	7	0.011	0.046	0.089	0.151
	9	0.009	0.038	0.074	0.126
$t = \frac{251}{252}T$					
γ	5	$*$	$*$	$*$	$*$
	7	$*$	$*$	$*$	$*$
	9	$*$	$*$	$*$	$*$

$T = 10$, the welfare cost increases from 2.245% to 3.434% as γ decreases from 9 to 5 at time $t = 0$; for $T = 20$, the welfare cost increases from 13.469% to 21.474%.

To summarize, calibrating our model with real data on stock returns and moving average signals, we find that technical investors with longer horizons would bear a substantial opportunity cost if they ignored estimation risk.

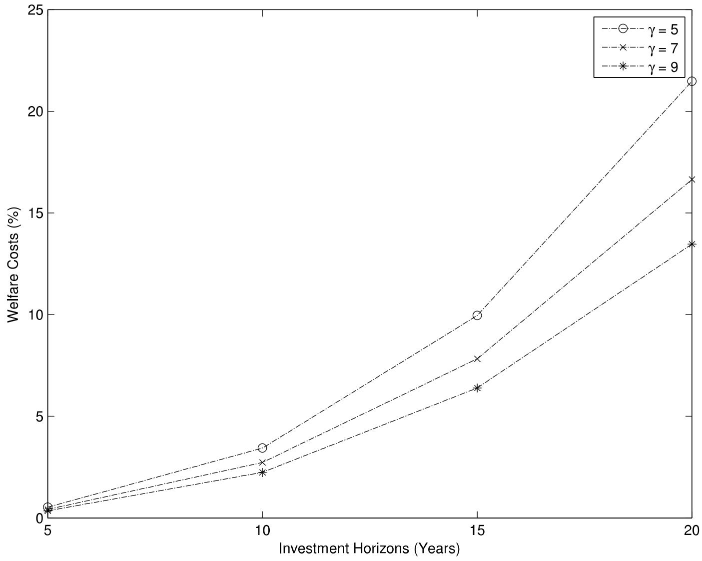

Figure 2.3. Welfare cost for ignoring estimation risk at time zero. This figure plots the welfare cost ∇_t as a function of the investment horizon T and the risk-aversion parameter γ at time $t = 0$. The results show that the welfare cost increases with horizon at an increasing rate. Besides, the welfare cost increases as the risk-aversion parameter decreases and the increase is stronger for longer horizons. The figures are measured in percent.

2.6 Conclusion

We study an investor's portfolio choice problem in which he uses a linear prediction model to forecast stock returns with moving average signals. The predictive power of the signals is uncertain and is formulated as an unknown slope parameter. The investor follows a simple Bayesian approach to account for estimation risk and optimally allocates his wealth to the stock.

This article contributes to the analysis and understanding of portfolio choice with technical analysis. First, we develop a model to examine the effects of uncertain predictive power of moving average signals on portfolio choice. Second, we derive an approximate solution for the optimal allocation to stocks. Third, we develop a simple numerical procedure to account for and decompose estimation risk. Fourth, calibrating the model with CRSP index data, we show that shorter-horizon investors bear little utility loss even if they ignore estimation risk. By contrast, such utility loss is sizable for investors with longer horizons.

2.7 Appendix

Appendix 2.A. Accounting for the Correlation between θ_t and $B_T - B_t$

Let ϱ_t denote the conditional correlation of θ_t and $B_T - B_t$, *i.e.*,

$$\varrho_t \equiv \frac{\mathrm{Cov}[\theta_t, B_T - B_t|\mathcal{F}_t]}{\sqrt{\mathrm{Var}[\theta_t|\mathcal{F}_t]}\sqrt{\mathrm{Var}[B_T - B_t|F_t]}} = \frac{\kappa_t}{\sqrt{v_t(T-t)}},$$

where $\kappa_t \equiv \mathrm{Cov}[\theta_t, B_T - B_t|\mathcal{F}_t]$. For simplicity, assuming that $(\theta_t, B_T - B_t)$ follows a bivariate Gaussian distribution, then we have

$$\mathrm{E}[B_T - B_t|\mathcal{F}_t, \theta_t] = \sqrt{\frac{T-t}{v_t}}\varrho_t(\theta_t - m_t),$$

and

$$\mathrm{Var}[B_T - B_t|\mathcal{F}_t, \theta_t] = (1 - \varrho_t^2)(T - t).$$

By log-normality, it can be shown that the inner expectation of (2.1) becomes

$$\begin{aligned} &\mathrm{E}[U(W_T)|\mathcal{F}_t, \theta_t] \\ &\qquad \propto \exp\left\{(1-\gamma)\xi[\theta_t - r + (\theta_t - m_t)\sigma\kappa_t - \tfrac{\xi}{2}(\gamma - (\gamma - 1)\varrho_t^2)\sigma^2](T - t)\right\}. \end{aligned}$$

Integrating out θ_t, after some simplification, the expected utility (2.1) becomes

$$\begin{aligned} &\mathrm{E}[U(W_T)|\mathcal{F}_t] \\ &\qquad \propto \exp\left\{(1-\gamma)\xi\{(m_t - r) - \tfrac{\xi}{2}[\gamma\sigma^2 + (\gamma - 1)(v_t(T - t) + 2\sigma\kappa_t)]\}(T - t)\right\}. \end{aligned}$$

It follows that the optimal allocation is

$$\xi_t^* = \frac{m_t - r}{\gamma\sigma^2 + (\gamma - 1)v_t(T - t) + 2(\gamma - 1)\sigma\kappa_t}.$$

One can proceed to use an Euler difference scheme as in Section 2.3.3 to simulate κ_t. However, it is also possible to analytically obtain bounds on κ_t, which can be useful to formulate a rough approximation of κ_t. First, by Cauchy-Schwarz inequality, we know that

$$|\kappa_t|^2 \leq \sqrt{\mathrm{Var}[\theta_t|\mathcal{F}_t]}\sqrt{\mathrm{Var}[B_T - B_t|\mathcal{F}_t]}$$

and so we have

$$-\sqrt{v_t(T-t)} \leq \kappa_t \leq \sqrt{v_t(T-t)}.$$

This implies that $\kappa_t \to 0$ as $t \to T$, meaning that the covariance will eventually become unimportant in determining ξ_t^*. Next, we attempt to refine the bounds by identifying the sign of κ_t. Recall that $\theta_t \equiv \beta_0 + \beta_1 \overline{\mathcal{X}}_t^T$. For the moment, let us assume that β_1 is known with $\beta_1 \geq 0$. Thus, it remains to examine the sign of

$$\mathrm{Cov}[\overline{\mathcal{X}}_t^T, B_T - B_t | \mathcal{F}_t] = \mathrm{E}[\overline{\mathcal{X}}_t^T (B_T - B_t) | \mathcal{F}_t]. \tag{2.12}$$

Let us approximate the stochastic processes $\{(P_\tau, D_\tau^{S,L}, X_\tau)\}_{\tau=t}^T$ with the following Euler difference scheme:

$$P_{t_{j+1}} = P_{t_j} + P_{t_j}[(\beta_0 + \beta_1 X_{t_j})\Delta t + \sigma(B_{t_{j+1}} - B_{t_j})],$$

$$D_{t_{j+1}}^{S,L} = D_{t_j}^{S,L} + \left[\frac{1}{s}(P_{t_j} - P_{t_{j-s}}) - \frac{1}{\ell}(P_{t_j} - P_{t_{j-\ell}})\right], \quad 0 < s < \ell,$$

$$X_{t_{j+1}} = \begin{cases} 1 & \text{if } D_{t_{j+1}}^{S,L} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 0, 1, \ldots, n-1$ with stepsize $\Delta t = T/n$ such that $t_0 = t$ and $T = t_n$. For a sufficiently large n, we can then approximate (2.12) with

$$\begin{aligned} \frac{1}{T-t}\int_t^T \mathrm{E}[X_\tau(B_T - B_t)|\mathcal{F}_t]\mathrm{d}\tau &\approx \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{E}[X_{t_j}(B_{t_n} - B_{t_0})|\mathcal{F}_{t_0}] \\ &= \frac{1}{n}\sum_{j=1}^{n-1} \mathrm{E}[X_{t_j}(B_{t_n} - B_{t_0})|\mathcal{F}_{t_0}] \end{aligned} \tag{2.13}$$

because $\mathrm{E}[X_{t_0}(B_{t_n} - B_{t_0})|\mathcal{F}_{t_0}] = X_{t_0}\mathrm{E}[B_{t_n} - B_{t_0}|\mathcal{F}_{t_0}] = 0$. Referring to the Euler difference scheme, we observe that X_{t_j} is an increasing (or nondecreasing) function of $B_{t_{j-1}}$ for $j = 1, \ldots, n-1$. Thus, we can view $X_{t_j} = \psi(B_{t_{j-1}})$ for an increasing function ψ. By Lemma 1(iii) in Lehmann (1966), it follows that

$$\mathrm{E}[X_{t_j}(B_{t_n} - B_{t_0})|\mathcal{F}_{t_0}] = \mathrm{E}[\psi(B_{t_{j-1}})(B_{t_n} - B_{t_0})|\mathcal{F}_{t_0}] \geq 0 \tag{2.14}$$

because $\mathrm{E}[B_{t_{j-1}}(B_{t_n} - B_{t_0})|\mathcal{F}_{t_0}] \geq 0$ for $j = 1, \ldots, n-1$. This implies that

$$\mathrm{Cov}[\overline{\mathcal{X}}_t^T, B_T - B_t|\mathcal{F}_t] \geq 0. \tag{2.15}$$

Hence, we can write the bounds on κ_t as

$$0 \leq \kappa_t \leq \sqrt{v_t(T-t)}. \tag{2.16}$$

Now we relax the assumption that β_1 is known and proceed as follows. Using the notation $\beta \equiv (\beta_0, \beta_1)$, by the law of total covariance, we have

$$\begin{aligned}
\kappa_t &\equiv \mathrm{Cov}[\theta_t, B_T - B_t|\mathcal{F}_t] \\
&= \mathrm{E}^{\beta}\{\mathrm{Cov}[\theta_t, B_T - B_t|\mathcal{F}_t, \beta]|\mathcal{F}_t\} + \mathrm{Cov}^{\beta}\{\mathrm{E}[\theta_t|\mathcal{F}_t, \beta], \mathrm{E}[B_T - B_t|\mathcal{F}_t, \beta]|\mathcal{F}_t\} \\
&= \mathrm{E}^{\beta}\{\mathrm{Cov}[\theta_t, B_T - B_t|\mathcal{F}_t, \beta]|\mathcal{F}_t\} + 0 \\
&= \mathrm{E}^{\beta}\{\beta_1 \mathrm{Cov}[\overline{\mathcal{X}}_t^T, B_T - B_t|\mathcal{F}_t, \beta]|\mathcal{F}_t\},
\end{aligned}$$

where the operators $\mathrm{E}^{\beta}\{\cdot\}$ and $\mathrm{Cov}^{\beta}\{\cdot\}$ indicate that the expectation and the covariance are taken over the distribution of β. Referring to the Euler scheme again, we observe that for all β_1'' and β_1' such that $\beta_1'' \geq \beta_1'$ we have $\psi(B_{t_{j-1}}; \beta_1'') \geq \psi(B_{t_{j-1}}; \beta_1')$. It follows from (2.14)-(2.15) that $\mathrm{Cov}[\overline{\mathcal{X}}_t^T, B_T - B_t|\mathcal{F}_t, \beta]$ is an increasing function of β_1, say $\varphi(\beta_1)$. Therefore, by Lemma 1(iii) in Lehmann (1966) again, we can conclude that

$$\kappa_t = \mathrm{E}^{\beta}\{\beta_1 \mathrm{Cov}[\overline{\mathcal{X}}_t^T, B_T - B_t|\mathcal{F}_t, \beta]|\mathcal{F}_t\} \quad = \mathrm{E}[\beta_1 \varphi(\beta_1)] \geq 0.$$

Thus, our previous bounds on κ_t, given by (2.16), also hold in this case.

Appendix 2.B. Proofs of Propositions

Proof of Proposition 2.1: This is a direct application of Theorem 12.7 in Liptser and Shiryaev (2001, p.36). ■

Proof of Proposition 2.2: Equation (4) is the second-order Taylor series expansion of $m_t = \mathrm{E}[\beta_0|\mathcal{F}_t]+\mathrm{E}[\beta_1\mathcal{E}_t^{\overline{\mathcal{X}}}(\beta)|\mathcal{F}_t]$ around the point $(m_t^{\beta_0}, m_t^{\beta_1})$. To obtain equation (5), we find the second-order and the first-order Taylor series expansions of the first and the second terms of

$$v_t = \mathrm{E}[\beta_1^2\mathcal{V}_t^{\overline{\mathcal{X}}}(\beta)|\mathcal{F}_t] + \mathrm{Var}[\beta_0 + \beta_1\mathcal{E}_t^{\overline{\mathcal{X}}}(\beta)|\mathcal{F}_t],$$

respectively. ■

Appendix 2.C. Formulas for Finite Difference Approximation

We use the following formulas to approximate the required partial derivatives in Proposition 2.2:

$$
\begin{aligned}
\partial_0 \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^\beta) &\approx \frac{\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}+\delta, m_t^{\beta_1}) - \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}-\delta, m_t^{\beta_1})}{2\delta}, \\
\partial_1 \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^\beta) &\approx \frac{\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}+\delta) - \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}-\delta)}{2\delta}, \\
\partial_{0,0} \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^\beta) &\approx \frac{\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}+\delta, m_t^{\beta_1}) - 2\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}) + \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}-\delta, m_t^{\beta_1})}{\delta^2}, \\
\partial_{1,1} \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^\beta) &\approx \frac{\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}+\delta) - 2\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}) + \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}-\delta)}{\delta^2}, \\
\partial_{0,1} \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^\beta) &\approx \frac{\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}+\delta, m_t^{\beta_1}+\delta) - \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}+\delta, m_t^{\beta_1}) - \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}+\delta)}{2\delta^2} \\
&\quad - \frac{\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}-\delta, m_t^{\beta_1}) + \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1}-\delta) - \mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}-\delta, m_t^{\beta_1}-\delta)}{2\delta^2} \\
&\quad + \frac{2\mathcal{E}_t^{\overline{\mathcal{X}}}(m_t^{\beta_0}, m_t^{\beta_1})}{2\delta},
\end{aligned}
$$

where we set $\delta = 0.01$. The formulas for $\mathcal{V}_t^{\overline{\mathcal{X}}}$ follow above accordingly.

2.8 References

Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. Journal of Financial Economics, 51(2), 245-271.

Avramov, D. (2002). Stock return predictability and model uncertainty. Journal of Financial Economics, 64(3), 423-458.

Bajgrowicz, P., & Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics, 106(3), 473-491.

Barberis, N. (2000). Investing for the long run when returns are predictable. Journal of Finance, 55(1), 225-264.

Bawa, V. S., Brown, S. J., & Klein, R. W. (1979). Estimation Risk and Optimal Portfolio Choice (North-Holland, New York).

Bossaerts, P., & Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies, 12(2), 405-428.

Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. The Journal of Finance, 54(5), 1609-1645.

Brandt, M. W., Goyal, A., Santa-Clara, P., & Stroud, J. R. (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. Review of Financial Studies, 18(3), 831-873.

Brennan, M. J. (1998). The role of learning in dynamic portfolio decisions. European Finance Review, 1(3), 295-306.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

Brown, D. P., & Jennings, R. H. (1989). On technical analysis. Review of Financial Studies, 2(4), 527-551.

Campbell, J. Y., & Viceira, L. M. (1999). Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics, 114(2), 433-495.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. Review of Financial Studies, 21(4), 1533-1575.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. Journal of Political Economy, 98(4), 703-738.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383-417.

Fama, E. F., & Blume, M. E. (1966). Filter rules and stock-market trading. Journal of Business, 39(1), 226-241.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis, 48(05), 1433-1461.

Jacquier, E., Kane, A., & Marcus, A. J. (2005). Optimal estimation of the risk premium for the long run and asset allocation: A case of compounded estimation risk. Journal of Financial Econometrics, 3(1), 37-55.

Jensen, M. C., & Benington, G. A. (1970). Random walks and technical theories: Some additional evidence. Journal of Finance, 25(2), 469-482.

Kandel, S., & Stambaugh, R. F. (1996). On the predictability of stock returns: An asset-allocation perspective. Journal of Finance, 51(2), 385-424.

Lehmann, E. L. (1966). Some concepts of dependence. The Annals of Mathematical Statistics, 1137-1153.

Liptser, R., & Shiryaev, A. N. (2001). Statistics of Random Processes: II. Applications (Springer-Verlag, New York).

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. Journal of Finance, 55(4), 1705-1765.

Lundtofte, F. (2008). Expected life-time utility and hedging demands in a partially observable economy. European Economic Review, 52(6), 1072-1096.

Menkhoff, L. (2010). The use of technical analysis by fund managers: International evidence. Journal of Banking and Finance, 34(11), 2573-2586.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60(7), 1772-1791.

Ready, M. J. (2002). Profits from technical trading rules. Financial Management, 31(3), 43-61.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. Journal of Economic Perspectives, 4(2), 19-33.

Stambaugh, R. F. (1999). Predictive regressions. Journal of Financial Economics, 54(3), 375-421.

Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance, 54(5), 1647-1691.

Wang, J. (1993). A model of intertemporal asset prices under asymmetric information. Review of Economic Studies, 60(2), 249-282.

Working, H. (1958). A theory of anticipatory prices. American Economic Review, 48(2), 188-199.

Xia, Y. (2001). Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. Journal of Finance, 56(1), 205-246.

Zhou, G., Zhu, Y., & Qiang, S. (2012). Asset allocation: can technical analysis add value?. International Journal of Portfolio Analysis and Management, 1(1), 43-58.

Zhu, Y., & Zhou, G. (2009). Technical analysis: An asset allocation perspective on the use of moving averages. Journal of Financial Economics, 92(3), 519- 544.

Chapter 3

Predictive Power of Technical Indicators and its Implications for Portfolio Choice

This article is a theoretical examination of the usefulness of technical forecasting for portfolio choice. Assuming stock returns follow a state-space model that allows for serial correlations, we show that moving average (MA) indicators have predictive power for returns in the context of linear prediction models. Calibrating our model with S&P 500 index and dividend yield data, we find that MA-based market timing can substantially improve annualized expected holding period returns. Our model also implies that shorter-horizon investors optimally time the market more aggressively and their portfolio profitability is more robust to parameter estimation errors in their return prediction models.

3.1 Introduction

In this article, we develop a simple model to theoretically study the usefulness of technical forecasting for portfolio choice. Assuming stock returns follow a state-space model that allows for serial correlations, we show that moving average (MA) indicators have predictive power for returns in the context of linear prediction models and thus are useful for investment. We then calibrate the model with real data and find that MA-based market timing can substantially improve annualized expected holding period returns over strategies that ignore time-varying investment opportunities.

This article is motivated by two observations. First, it is now generally accepted that stock returns are predictable and thus this admits potentially useful technical indicators because they may be correlated with the time-varying drift of returns. By contrast, earlier theoretical studies often assume that stock price increments are independent, which completely rules out any use of technical forecasting. In this article, we take stock return predictability as a given fact and study its implications for the usefulness of technical analysis.

Second, previous empirical studies may not correctly reflect the true profitability of technical analysis. This is because most of these studies assume that technical investors use a naive "all-or-nothing" strategy, that is allocating 100% of their wealth into stocks when they observe a buy signal but nothing otherwise. In this article, we allow the allocation of wealth to vary continuously with the MA indicators.

Our modelling approach is as follows. We consider a two-asset economy with a risk-free asset and a risky stock. We assume that the drift of stock returns follows the Ornstein-Uhlenbeck process but investors do not observe this drift and they do not have any knowledge about the stochastic process determining the drift. That is, the true stock return model remains unknown to the investors. Investors observe a MA indicator denoted by X_t, measuring the distance between the current price and the moving average of past prices, and conjecture a linear predictor $\alpha + \beta X_t$ for returns. The "best" values of (α, β) are defined as the parameter values that minimize the mean integrated squared (forecast) error over a given forecast horizon. The technical indicator X_t is said to have predictive power if such β is nonzero. In our case, X_t is a continuous variable, rather than a zero-one variable, as is commonly studied in the literature.

Next, we construct and examine the expected profitability of the optimal portfolio strategy of a log-utility investor, assuming he knows the "best" model parameters. Then, we consider the case when the investor fails to identify such "best" parameter values. Finally, we calibrate the model with S&P 500 stock index and dividend yield data to calculate the annualized expected holding period returns (HPRs), as a measure of expected profitability, of such optimal MA-based strategies.

We find that our optimal strategies substantially improve annualized HPRs for a range of reasonable investment horizons. Our model also implies that shorter-horizon investors optimally time the market more aggressively and their portfolio profitability is more robust to estimation errors in their return prediction models.

To the best of our knowledge, this is the first study to theoretically justify the predictive powers of MA indicators, an important class of technical indicators, on stock returns in the context of linear predictors and to discuss the implications of such predictive powers for portfolio choice.

Our study offers a number of useful insights about technical analysis. First, we propose a simple measure of predictive power that explicitly takes into account expected forecast errors over a given horizon in continuous time. This measure represents expected gain in forecast accuracy using conditional forecasts of the form $\alpha + \beta X_t$ over the long-run mean of returns alone. Second, we relate this measure of predictive power to expected HPRs implied by the MA-based strategies. Third, we provide comparative statics to examine how some key model parameters of the true but unknown stock return model affect the profitability of the optimal MA-based strategies.

Fourth, although this article is closely related to the work of Zhu and Zhou (2009), there are important differences. They show that, when stock returns are predictable, some market timing strategies based on zero-one MA signals (buy or sell) improve investors' expected utility relative to strategies ignoring time-varying investment opportunities. Our modelling approach differs from theirs in that we explicitly use a linear prediction model to study the predictive power of continuous MA indicators. This approach allows us to understand better the statistical foundations of technical forecasting, which attract relatively little attention in previous studies as mentioned by Neftci (1991). This article also builds upon the literature on applied portfolio modelling. It is common to incorporate predictability in returns to study how such predictability affects investors' optimal portfolio choice (see, *e.g.* Kandel and Stambaugh, 1996; Brennan, Schwartz, and Lagnado, 1997; Barberis, 2002; Xia, 2001). However, in these studies, the linear prediction model is (implictly) assumed to be correctly specified, *i.e.*, the form of the prediction model coincides with the true return model except with unknown parameters to be empirically estimated. We deviate from this assumption and view the linear prediction model as a statistical approximation of the true return model. Hence, we need to define the "best" model parameters in an appropriate statistical sense such that we can view the *misspecified* prediction model as a useful approximation.

The treatment of this chapter is closely related to the previous one. We will once again consider an individual investor who trades continuously in a two-asset economy. The key difference is that we now shift from a Bayesian perspective to a frequentist perspective when measuring the predictive power of the moving averages. Such a shift in perspective is nontrivial because the results from a Bayesian perspective are by nature data-dependent and thus fall short in offering a more general probabilistic foundation. Indeed, the Bayesian investor simply proposes a linear prediction model without knowing what his estimates of the predictive parameters will converge to, or whether the estimates will converge at all. Such concerns are irrelevant for a Bayesian investor as he believes that the data speak for themselves eventually. For example, the slope estimate may converge to zero over a long period, then he will conclude that his prediction model is not useful but before so the model is still considered "useful". However, for a frequentist investor, he considers the effectiveness of his prediction model across infinite hypothetical trails *before* data (or very limited data) are collected.

The article is organized as follows. The next section discusses some issues about measuring predictive power. Section 3.2 introduces our model and provides analytic results. Section 3.3 describes the data and calibrates the model. Section 3.4 presents numerical results. Section 3.5 summarizes and concludes the article.

3.2 Background

3.2.1 Technical Analysis and Stock Return Predictability

Technical analysts believe that technical analysis can generate useful indicators to help predict stock returns. A straightforward approach to examine the predictive powers of technical indicators is via the linear prediction model, $R_{t+1} = \alpha + \beta X_t + \varepsilon_t$, where R_{t+1} is the return of a stock form period t to $t+1$, X_t is some technical indicator observed at period t. Researchers then use a data sample to test the null hypothesis that $\beta = 0$. The technical indicator is said to have predictive power if the statistical test rejects the null hypothesis at some conventional significance level.

Remarkably, the empirical evidence on the usefulness of technical analysis in the literature is mixed and inconclusive. While some studies find strong evidence in favor of the usefulness of technical trading (see, *e.g.* Brock, Lakonishok, and LeBaron, 1992; Lo, Mamaysky, and Wang, 2000; Zhou, Zhu, and Qiang, 2012; Han, Yang, and Zhou, 2013), many other studies provide evidence questioning its usefulness (see, *e.g.* Allen and Karjalainen, 1999; Sullivan, Timmermann, and White, 1999; White, 2000; and Bajgrowicz and Scillet, 2012). Since typically only a single time series for a given phenomenon of interest is available, it is challenging to test the true effectiveness of any technical trading rule in practice. Theoretical studies, like this article, can therefore provide complementary insights for this empirical debate.

By contrast, it is now commonly agreed that stock returns are predictable in the sense that future stock returns are correlated with the current values of some observable predictive variables (also called state variables). There is ample evidence in favor of return predictability. The most powerful predictive variables include the dividend yield or the net payout yield (Campbell and Shiller, 1988; Boudoukh et al., 2007), the earnings-price ratio (Campbell and Shiller, 1988), the book-to-market ratio (Kothari and Shanken, 1997), and the short-term interest rate (Ang and Bekaert, 2007). Intuitively, such evidence does allow for the predictive power of some technical indicators because these indicators may be correlated with these predictive variables. This article attempts to provide such a theoretical basis.

3.2.2 Measuring Predictive Power

It is useful to elaborate on some issues about measuring predictive power proposed by Diebold and Kilian (2001) in the context of linear prediction models.

First, the predictive power of a predictive variable should be measured relative to a benchmark because it is always a matter of degree. The simplest benchmark is the long-run (or unconditional) mean of the dependent variable, provided that it exists. We can then relate the predictive power to the slope parameter in the sense that the predictive variable provides *additional* information (over the long-run mean) for forecasting if and only if the slope is nonzero. However, the value of the slope does not directly measure how useful the predictive variable is because its value depends on the scale of the predictive variable. A better measure is therefore called for.

Second, predictive power should be measured over a relevant forecast horizon and loss function. The difference between the expected losses of forecast by using the best linear predictor to that of the long-run mean can be interpreted as a measure of predictive power. For example, if the difference in expected loss of forecast is larger for shorter horizons relative to longer horizons (*i.e.*, a gain in expected forecast accuracy) then we say that the predictive variable has stronger predictive power at shorter horizons.

Third, predictive power is a *population* concept, not a property of any particular sample path. While we can estimate the predictive power from a sample path, it is important to use an appropriate econometric approach. This issue is related to a common critique in the literature that one can always find a "useful" trading rule by an extensive rule searching. If the "best" trading rule is picked by pure chance rather than any inherent merit of this rule, then its observed predictive power is a positively biased estimate of its future predictive power. Such bias is commonly called *data-snooping bias* (see, *e.g.* Lo and MacKinlay, 1990). The literature proposes several econometric approaches to alleviate data-snooping bias such as generic algorithms (Allen and Karjalainen, 1999), bootstraps (Sullivan, Timmermann, and White, 1999; White, 2000), and false discovery rates (Bajgrowicz and Scaillet, 2012). Still, the quality of estimation and statistical inference depend highly on the availability of suitable data.

This article takes a new perspective. We deviate from the statistical testing approach and propose a simple theoretical model that allows us to measure the predictive powers of technical indicators (specifically, the MA indicators) and hence the expected profitability of market timing based on these indicators. This model approach gives us useful insights to understand better the economic relevance of technical analysis in practice.

3.3 The Model and Analytic Results

3.3.1 The Basic Setting

Consider an investor with a long horizon who trades continuously in a two-asset economy in which a risk-free asset pays an instantaneous rate of interest r, and a risky stock represents the aggregate equity market. We fix a finite horizon $[0, T]$, where T is measured in years. The cum-dividend stock price P_t grows according to the following process

$$\frac{\mathrm{d}P_t}{P_t} = \mu_t \mathrm{d}t + \sigma \mathrm{d}B_t, \tag{3.1}$$

where the percentage volatility σ is known to the investor due to observable quadratic variations of P_t; B_t is a Brownian motion defined on the probability space $(\Omega, \mathbb{P}, \mathcal{F})$ with a standard filtration $\mathcal{F} \equiv \{\mathcal{F}_t : t \leq T\}$; the percentage drift, $\mu_t \in \mathcal{F}_t$, grows according to the Ornstein-Uhlenbeck process

$$\mathrm{d}\mu_t = \lambda(\overline{\mu} - \mu_t)\mathrm{d}t + \eta\, \mathrm{d}Z_t, \tag{3.2}$$

with $\mu_0 = \overline{\mu}$, where λ, $\overline{\mu}$, and η are constant parameters; Z_t is a $\mathbb{P}$-Brownian motion correlated with B_t with correlation coefficient ρ, *i.e.*, $\mathrm{d}B_t \mathrm{d}Z_t = \rho\, \mathrm{d}t$. Equations (3.1)-(3.2) are commonly referred to as a *state-space model*. We assume that the investor knows the long-run mean of stock returns, $\overline{\mu}$, but does not observe the drift, μ_t. Furthermore, he has no knowledge about the stochastic process determining μ_t as described by (3.2), including ρ. However, he observes a technical indicator X_t, a function of past stock prices, that is potentially useful to estimate μ_t, or equivalently, to forecast returns $\mathrm{d}P_t/P_t$.

Let L be a positive parameter. We consider

$$X_t \equiv \log P_t - \log A_t, \tag{3.3}$$

where

$$A_t \equiv \exp\left\{L \int_{-\infty}^{t} e^{-L(t-\tau)} \log P_\tau \mathrm{d}\tau\right\} \tag{3.4}$$

is an exponentially weighted moving average (MA) of stock prices with L controlling the window size. We interpret X_t as the percentage difference of the current stock price and the moving average at time t and call it a *moving average indicator.* Note that if $P_t = c$, where c is a positive constant, then $A_t = c$. We assume that the current stock price and the history, $\{P_t : t \leq 0\}$, is known such that $(P_0, A_0) \in \mathcal{F}_0$ are taken as given parameters. We use the shorthand notation $p_t \equiv \log P_t$ and $a_t \equiv \log A_t$ such that $X_t = p_t - a_t$.

Lemma 3.1. *The moving average indicator X_t defined by* (3.3)-(3.4) *has a long-run mean given by*

$$\overline{X} \equiv \lim_{t \to \infty} \mathrm{E}[X_t] = \frac{2\overline{\mu} - \sigma^2}{2L},$$

where we use the shorthand $\mathrm{E}[\cdot] \equiv \mathrm{E}[\cdot|\mathcal{F}_0]$.

Proof. Applying Itô's rule to obtain the solution of p_t and taking expectation, we have

$$\mathrm{E}[p_t] = p_0 + (\overline{\mu} - \tfrac{\sigma^2}{2})t, \tag{3.5}$$

where we use the fact that $\mathrm{E}[\mu_t] = \overline{\mu}$ (because we set $\mu_0 = \overline{\mu}$). Next, we find $\mathrm{E}[a_t]$

by using

$$\begin{aligned}\mathrm{E}[a_t] &= L\int_{-\infty}^{t} e^{-L(t-\tau)}\mathrm{E}[p_\tau]\mathrm{d}\tau \\ &= a_0 e^{-Lt} + L\int_{0}^{t} e^{-L(t-\tau)}\mathrm{E}[p_\tau]\mathrm{d}\tau.\end{aligned} \tag{3.6}$$

Substituting (3.5) into (3.6) and evaluating the integral, we have

$$\mathrm{E}[a_t] = a_0 e^{-Lt} + p_0(1-e^{-Lt}) + (\overline{\mu} - \tfrac{\sigma^2}{2})\left(t - \frac{1-e^{-Lt}}{L}\right). \tag{3.7}$$

Thus, subtracting (3.7) from (3.5), we have

$$\mathrm{E}[X_t] = X_0 e^{-Lt} + (\overline{\mu} - \tfrac{\sigma^2}{2})\left(\frac{1-e^{-Lt}}{L}\right).$$

Taking the limit as $t \to \infty$, we obtain

$$\lim_{t\to\infty} \mathrm{E}[X_t] = \frac{2\overline{\mu} - \sigma^2}{2L},$$

as desired. ■

The assumption that $\overline{\mu}$ is known to the investor implies that $\overline{X}$ is also a known quantity. The role of $\overline{X}$ will become clear in subsequent sections.

We remind our readers that we always use the notation $\mathrm{E}[\cdot]$ to denote expectations conditional on the initial conditions in this article, not to be confused with the long-run (invariant) expectations, indicated by an upper bar, such as $\overline{\mu}$ and $\overline{X}$.

Let $\tilde{\mathbb{P}}$ denote the investor's subjective probability measure and $\tilde{\mathcal{F}}_t = \{P_\tau : \tau \leq t\}$, with $\tilde{\mathcal{F}}_t \subset \mathcal{F}_t$, denote the investor's information set at time t. The investor conjectures a *linear predictor* $\alpha + \beta X_t$ of μ_t, and thus under the reference model probability $\tilde{\mathbb{P}}$ the dynamics of stock price are described by the following *linear* *prediction model*

$$\frac{\mathrm{d}P_t}{P_t} = (\alpha + \beta X_t)\mathrm{d}t + \sigma \mathrm{d}\tilde{B}_t, \tag{3.8}$$

where (α, β) are model parameters to be determined; $\tilde{B}_t$ is a $\tilde{\mathbb{P}}$-Brownian motion adapted to the investor's information set $\tilde{\mathcal{F}}_t$. For simplicity, we follow Xia (2001) to set $\alpha \equiv \overline{\mu} - \beta\overline{X}$ so that α is known whenever the value of β is determined. We justify this short-cut approach in the Appendix. We can now rewrite the prediction model (3.8) as

$$\frac{\mathrm{d}P_t}{P_t} = (\overline{\mu} + \beta\hat{X}_t)\mathrm{d}t + \sigma \mathrm{d}\tilde{B}_t, \tag{3.9}$$

where we define $\hat{X}_t \equiv X_t - \overline{X}$. We note that $\hat{X}_t > 0$ (< 0) if and only if $p_t - a_t > \overline{X}$ $(< \overline{X})$. Thus, we can interpret $\overline{X}$ as a *bandwidth parameter* in the context of moving average rules. Since any arbitrary choice of the slope parameter β is unlikely to provide useful forecasts, we shall define the "best" value of β in an appropriate statistical sense. To move on to the investor's portfolio choice problem, let us first assume that the investor knows such β and we shall discuss some economic implications if he incorrectly identifies β.

Given an investment horizon T and an initial wealth W_0, the investor chooses a portfolio strategy $\xi \equiv \{\xi_t\}_{t=0}^{T}$ to maximize his expected utility of wealth,

$$\max_{\xi} \tilde{\mathrm{E}}[U(W_T)],$$

subject to wealth dynamics

$$\frac{\mathrm{d}W_t}{W_t} = r\mathrm{d}t + \xi_t(\overline{\mu} + \beta\hat{X}_t - r)\mathrm{d}t + \xi_t\sigma\mathrm{d}\tilde{B}_t,$$

where the operator $\tilde{\mathrm{E}}[\cdot] \equiv \tilde{\mathrm{E}}[\cdot|\tilde{\mathcal{F}}_0]$ indicates that the expectation is taken over the subjective probability measure $\tilde{\mathbb{P}}$ conditional on $\tilde{\mathcal{F}}_0$.

In this article, we assume the logarithmic utility function

$$U(W_T) = \log W_T.$$

In this case, the *optimal strategy* is given by

$$\xi_t^* = \frac{\overline{\mu} + \beta \hat{X}_t - r}{\sigma^2}. \tag{3.10}$$

3.3.2 Defining the "Best" Value for the Slope Parameter β

In most previous studies in portfolio choice, the value of the slope parameter β is estimated using real data. However, this approach is inappropriate for this article because the form of the investor's prediction model does not coincide with that of the true model. Thus, we cannot expect the standard estimators to have the usual asymptotic limit because the true value of β remains undefined. We now attempt to define β in such a way that we can view the investor's prediction model as a reasonable statistical approximation of the true model for stock returns.

Fixing a point time t in $[0, T]$ and an arbitrarily small $h > 0$, the h-period stock return under $\tilde{\mathbb{P}}$ is

$$R_{t,t+h} \equiv \frac{P_{t+h} - P_t}{P_t} = (\overline{\mu} + \beta \hat{X}_t)h + \sigma(\tilde{B}_{t+h} - \tilde{B}_t) + o(h),$$

with $o(h)/h \to 0$ as $h \to 0$. Thus, given observations $(P_t, \hat{X}_t)$ at time t, the conditional forecast of the h-period return is

$$\tilde{R}_{t,t+h} \equiv \tilde{\mathrm{E}}[R_{t,t+h} | \tilde{\mathcal{F}}_t] = (\overline{\mu} + \beta \hat{X}_t)h + o(h).$$

A common measure of forecast accuracy over a time interval $[0,T]$ is the *mean integrated squared error* (MISE). For future observations $\{R_{t,t+h}\}_{t=0}^{T}$ generated by the true model $\mathbb{P}$ and future conditional forecasts $\{\tilde{R}_{t,t+h}\}_{t=0}^{T}$, the MISE is defined as

$$\begin{aligned}\text{MISE} &= \mathrm{E}\int_0^T (R_{t,t+h}-\tilde{R}_{t,t+h})^2 \mathrm{d}t \\ &= \mathrm{E}\int_0^T \left[\frac{P_{t+h}-P_t}{P_t} - (\overline{\mu}+\beta\hat{X}_t)h - o(h)\right]^2 \mathrm{d}t,\end{aligned} \tag{3.11}$$

where the notation $\mathrm{E}[\cdot]$ indicates that the expectation is taken with respect to the true (joint) densities of $\{(P_t,\hat{X}_t)\}_{t=0}^{T}$, conditional on $\mathcal{F}_0$. We find the *best linear predictor* of the h-period returns by choosing β such that (3.11) is minimized.

Expanding the squared term in the last line of (3.11), we have

$$\mathrm{E}\int_0^T \left[(\beta^2\hat{X}_t^2 + 2\overline{\mu}\hat{X}_t)h^2 - 2\beta\hat{X}_t\frac{P_{t+h}-P_t}{P_t}h\right]\mathrm{d}t + \text{other terms}, \tag{3.12}$$

where "other terms"' include terms independent of β and the "$o(h)$" term. Since $h>0$, minimizing (3.12) is equivalent to minimizing

$$\mathrm{E}\int_0^T \left[\frac{1}{2}\beta^2\hat{X}_t^2 + \overline{\mu}\beta\hat{X}_t - \beta\hat{X}_t\frac{1}{h}\frac{P_{t+h}-P_t}{P_t}\right]\mathrm{d}t, \tag{3.13}$$

where we divide the last line of (3.12) by $2h^2$ and drop the "other terms". Expanding (3.13) gives us

$$\frac{1}{2}\beta^2\mathrm{E}\int_0^T \hat{X}_t^2\mathrm{d}t + \beta\mathrm{E}\int_0^T \hat{X}_t\overline{\mu}\,\mathrm{d}t - \beta\mathrm{E}\left[\frac{1}{h}\int_0^T \hat{X}_t\frac{P_{t+h}-P_t}{P_t}\mathrm{d}t\right]. \tag{3.14}$$

The term in the last expectation operator is the so-called *Russo-Vallois integral*. Since the investor is interested in forecasting instantaneous stock returns, we take the limit as $h \to 0$ to obtain

$$\mathrm{E}\left[\frac{1}{h}\int_0^T \hat{X}_t \frac{P_{t+h}-P_t}{P_t}\mathrm{d}t\right] \to \mathrm{E}\int_0^T \hat{X}_t \frac{\mathrm{d}P_t}{P_t} = \mathrm{E}\int_0^T \hat{X}_t \mu_t \mathrm{d}t, \tag{3.15}$$

where the equality follows from (3.1).

Substituting (3.15) in (3.14) and using the definition $\hat{\mu}_t \equiv \mu_t - \overline{\mu}$, we can replace the original minimization problem by

$$\min_\beta \mathcal{L}(\beta;T) \equiv \frac{1}{2}\beta^2 \mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t - \beta \mathrm{E}\int_0^T \hat{X}_t \hat{\mu}_t \mathrm{d}t, \tag{3.16}$$

where $\mathcal{L}$ is the *loss function* implied by the MISE criterion. It turns out that the negative of $\mathcal{L}$ is equivalent to the expected conditional log-likelihood of $\{P_t\}_{t=0}^T$ (see Appendix 3.A). Differentiating (3.16) with respect to β and equating it to zero, we obtain the best value of β,

$$\beta = \frac{\mathrm{E}\int_0^T \hat{X}_t \hat{\mu}_t \mathrm{d}t}{\mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t}. \tag{3.17}$$

Substituting β back to the loss function (3.16), with a slight abuse of notation, the minimized value of $\mathcal{L}$ is given by

$$\begin{aligned} \mathcal{L}(\beta;T) &= -\frac{1}{2}\frac{(\mathrm{E}\int_0^T \hat{X}_t \hat{\mu}_t \mathrm{d}t)^2}{\mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t} \\ &= -\frac{1}{2}\beta^2 \mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t \le 0. \end{aligned} \tag{3.18}$$

Since $\mathcal{L}(\beta;T)$ is bounded above by zero (if $\beta = 0$), we can view $|\mathcal{L}(\beta;T)|$ as a measure of *expected gain in forecast accuracy* using the best linear predictor $\overline{\mu} + \beta\hat{X}_t$ over the long-run mean $\overline{\mu}$ alone (which corresponds to $\mathcal{L}(0;T) = 0$).

Hence, when β is nonzero, we say that the technical indicator X_t has *predictive power* on stock returns in the sense that there is an expected gain in forecast accuracy. In general, β is nonzero unless stock returns come from an IID model. That is, when $\mu_t = \mu_0 = \overline{\mu}$ for all $t \in [0, T]$ (*i.e.*, $\lambda = \eta = 0$). In this case, we have $\hat{\mu}_t = 0$ for all $t \in [0, T]$ and hence $\beta = 0$. We note that this definition of predictive power is restrictive because we essentially assume that the investor fixes the same β for the entire horizon. A more realistic model would allow the investor to update the slope parameter. However, this would make the model much less tractable.

The results above may appear to be similar to that of linear projection in standard time-series econometrics. However, our definition of β does not require the processes to be asymptotically stationary. While the concept of MISE is not new, we are unaware of its use as a defining criterion for model parameters in continuous time.

In Appendix 3.B, we verify that if the investor used the original prediction model (3.8), then the MISE minimizing (α, β) would indeed be $\alpha = \overline{\mu} - \beta\overline{X}$ with β as given by (3.17), if we assume X_0 starts at its long-run mean (*i.e.*, $X_0 = \overline{X}$). It is useful to note that if $X_0 = \overline{X}$ then

$$\mathrm{E}\int_0^T \hat{X}_t \mathrm{d}t = \mathrm{E}\int_0^T (X_t - \overline{X})\mathrm{d}t = 0,$$

for any horizon T. This property will allow us to simplify our analytical study. In this article, we set $P_0 = 1$ such that $p_0 = 0$ and $X_0 = \overline{X}$ such that $a_0 = -\overline{X}$.

3.3.3 Computing the "Best" Slope Parameter β

We now evaluate the two expectations in β given by (3.17). Expanding and simplifying (3.17), we have that

$$\beta = \frac{\int_0^T \{\mathrm{E}[p_t\mu_t] - \mathrm{E}[a_t\mu_t] - \overline{\mu}(\mathrm{E}[p_t] - \mathrm{E}[a_t])\}\mathrm{d}t}{\int_0^T \{\mathrm{E}[p_t^2] + \mathrm{E}[a_t^2] - 2\mathrm{E}[p_t a_t] - 2\overline{X}(\mathrm{E}[p_t] - \mathrm{E}[a_t]) + \overline{X}^2\}\mathrm{d}t}.$$

To find the moment equations required to compute β, it is useful to write the dynamics of (p_t, a_t, μ_t) in matrix form. Let us define $y_t \equiv (p_t, a_t, \mu_t)^\top$. By noting that the dynamics of p_t and a_t are

$$\begin{aligned}
\mathrm{d}p_t &= (\overline{\mu} - \tfrac{\sigma^2}{2})\mathrm{d}t + \sigma \mathrm{d}B_t, \\
\mathrm{d}a_t &= L(p_t - La_t)\mathrm{d}t,
\end{aligned}$$

together with "$\mathrm{d}\mu_t$" given by (3.2), the dynamics of y_t can be compactly written as

$$\mathrm{d}y_t = (M_0 + M_1 y_t)\mathrm{d}t + \Sigma\, \mathrm{d}\vec{B}_t, \tag{3.19}$$

where $\vec{B}_t \equiv (B_{1t}, B_{2t})^\top$ is a vector of two independent Brownian motions (we can write $B_t = B_{1t}$ and $Z_t = \rho B_{1t} + \sqrt{1-\rho^2}B_{2t}$),

$$M_0 \equiv \begin{pmatrix} -\sigma^2/2 \\ 0 \\ \lambda\overline{\mu} \end{pmatrix}, \quad M_1 \equiv \begin{pmatrix} 0 & 0 & 1 \\ L & -L^2 & 0 \\ 0 & 0 & -\lambda \end{pmatrix}, \quad \Sigma \equiv \begin{pmatrix} \sigma & 0 \\ 0 & 0 \\ \eta\rho & \eta\sqrt{1-\rho^2} \end{pmatrix},$$

with given initial values $y_0 = (p_0, a_0, \mu_0)^\top$. To evaluate the expectations in β, we require the moment equations $m_t \equiv \mathrm{E}[y_t]$ and $v_t \equiv \mathrm{E}[y_t y_t^\top]$. We have already found m_t, given by (3.5), (3.7), and $\mathrm{E}[\mu_t] = \overline{\mu}$, so it remains to find v_t.

Applying Itô's product rule, we have

$$\begin{aligned}\mathrm{d}(y_t y_t^\top) &= y_t \mathrm{d}y_t^\top + (\mathrm{d}y_t)y_t^\top + (\mathrm{d}y_t)\mathrm{d}y_t^\top \\ &= y_t \mathrm{d}y_t^\top + (\mathrm{d}y_t)y_t^\top + \Sigma\Sigma^\top \mathrm{d}t,\end{aligned}$$

substituting "$\mathrm{d}y_t$" given by (3.19), and taking expectations, we obtain the system of ordinary differential equations

$$\frac{\mathrm{d}v_t}{\mathrm{d}t} = M_1 v_t + v_t M_1^\top + M_0 m_t^\top + m_t M_0^\top + \Sigma\Sigma^\top, \tag{3.20}$$

with the given initial values $v_0 = y_0 y_0^\top$. The solutions to (3.20) can be obtained by using a symbolic mathematical computation program. It turns out that while there are closed-form solutions for v_t, their algebraic solutions are too lengthy to display in this article. Instead, we first estimate the model parameters, namely $(\sigma, \lambda, \overline{\mu}, \eta, \rho)$, using real data in Section 3.4 and then we substitute the estimates into the solutions of v_t to illustrate their functional forms (for various window sizes L) in Section 3.5. After that, we evaluate β for different values of (L, T).

3.3.4 Stability Analysis of the Moments in β and Some Practical Implications

In this section, we study the stability properties of the moments in β. The discussion is relatively technical but is included in the main body of this article because such properties have important econometric and mathematical implications.

Let $\phi_t \equiv (m_t, \mathrm{vech}(v_t))^\top$ be a 9×1 vector of the first and the second moments of $y_t \equiv (p_t, a_t, \mu_t)^\top$, where $\mathrm{vech}(\cdot)$ denotes the half-vectorization operator (that is, $\mathrm{vech}(v_t)$ is a 6×1 vector by vectorizing only the lower triangular part of the symmetric 3×3 matrix v_t). Taking expectations of (3.19), we have

$$\frac{\mathrm{d}m_t}{\mathrm{d}t} = M_0 + M_1 m_t. \tag{3.21}$$

By equations (3.20) and (3.21), we can compactly write the dynamics of ϕ_t as

$$\frac{\mathrm{d}\phi_t}{\mathrm{d}t} = A_0 + A_1 \phi_t, \tag{3.22}$$

where A_0 is a 9×1 vector and A_1 is a 9×9 matrix given in Appendix 3.C.

Definitions. Consider the system of differential equations $\mathrm{d}\phi_t/\mathrm{d}t$. Given any initial value ϕ_0, we denote its solution by $\phi(t|\phi_0)$. Similar to Definitions 2.1 by Brock and Malliaris (1992), the solution is called *stable* (with respect to the initial conditions) if for every ε, there is a $\delta = \delta(\varepsilon)$ such that $|\phi_0 - \phi_0'| < \delta$ implies

$$|\phi(t|\phi_0) - \phi(t|\phi_0')| < \varepsilon$$

for all $t \geq 0$. The solution is *asymptotically stable* if it is stable and if $|\phi(t|\phi_0) - \phi(t|\phi_0')| \to 0$ as $t \to \infty$. The solution is *unstable* if it is not stable.

In words, if the solution $\phi(t|\phi_0)$ is stable, then for two sufficiently close initial values ϕ_0 and ϕ_0', with $\phi_0 \neq \phi_0'$, the solutions $\phi(t|\phi_0)$ and $\phi(t|\phi_0')$ remain very near to one another for $t \geq 0$. Furthermore, if the solution is asymptotically stable, then the stochastic process y_t has both long-run (invariant) first and second moments (*i.e.*, y_t is *asymptotically covariance-stationary*). If the solution $\phi(t|\phi_0)$ is unstable, then $\phi(t|\phi_0)$ and $\phi(t|\phi_0')$ drift away from one another as $t \to \infty$, even if ϕ_0 and ϕ_0' are close.

Following Theorem 4.1 by Brock and Malliaris (1992), the stability properties of (3.22) are given in terms of the eigenvalues of A_1. In particular, we have the following summary to check stability.

(i) The solution $\phi(t|\phi_0)$ is stable if all eigenvalues of A_1 have nonpositive real parts and if every eigenvalue of A_1 which as a zero real part is a simple zero of the characteristic polynomial of A_1.

(ii) The solution $\phi(t|\phi_0)$ is asymptotically stable if all eigenvalues of A_1 have negative real parts.

(iii) The solution $\phi(t|\phi_0)$ is unstable if at least one eigenvalue of A_1 has positive real part.

We find that the nine eigenvalues of A_1 are:

$$0, 0, -L^2, -L^2, -2L^2, -(L^2 + \lambda), -\lambda, -\lambda, \text{and} - 2\lambda.$$

Since there are two nonnegative eigenvalues (two zeros), the solution $\phi(t|\phi_0)$ is not asymptotically stable, meaning that y_t is not an asymptotically covariance-stationary process. This finding has important implications for econometric estimation of β. In particular, the standard OLS estimator based on a single time series of stock price will *not* converge to β even if the sample size goes to infinity and so better estimation strategy is required. Although it is beyond the scope of this article to discuss about estimation methodology in details, we note that it is possible to estimate β by resampling approaches. For example, we can block bootstrap stock returns to construct artificial sample paths of stock price. For each block bootstrap sample, we run a regression to obtain an OLS estimate of β. The average of these OLS estimates can then be taken as a final estimate of β.

We next check whether the solution $\phi(t|\phi_0)$ is at least stable with respect to the initial conditions. In a theoretical perspective, the stability properties are desirable because it implies that our numerical results based on computed β (presented in Section 3.5) are robust to our arbitrary choice of initial values (setting $p_0 = 0$, $a_0 = -\overline{X}$, and $\mu_0 = \overline{\mu}$) in the sense that we can reasonably expect similar quantitative results even if we deviate slightly from these initial values.

To check that $\phi(t|\phi_0)$ is stable, it suffices to show that the two eigenvectors corresponding to the two eigenvalue zeros are linearly independent (if so, we say the eigenvalue zeros are *simple zeros*). We find that the two eigenvectors are $(0, 0, 0, L^2, L, 0, 1, 0, 0)^\top$ and $(0, 0, 0, 0, L/2, 0, 1, 0, 0)^\top$, which are therefore linearly independent. Thus, we can conclude that $\phi(t|\phi_0)$ is stable with respect to the initial conditions.

We now finish our technical discussion about the stability properties of the moments in β and return to our portfolio choice problem to obtain some economic insights from the model.

3.3.5 Measuring Economic Value of Market Timing

The optimal portfolio allocation ξ_t^* depends on the current value of the MA indicator X_t augmented with its long-run mean, namely $\hat{X}_t$. We call this dependence *market timing*. Let us define the *IID strategy* by

$$\xi_t^o \equiv \frac{\overline{\mu} - r}{\sigma^2}, \tag{3.23}$$

which would be used by a log-utility investor who disregards return predictability (*i.e.*, assuming returns following an IID model with $\beta = 0$).

Note that we can write

$$\xi_t^* = \xi_t^o + \frac{\beta \hat{X}_t}{\sigma^2}, \tag{3.24}$$

implying

$$\frac{\partial \xi_t^*}{\partial \hat{X}_t} = \frac{\beta}{\sigma^2}.$$

Thus, β measures the change in optimal allocation to stocks with respect to a small change in the current value of the MA indicator. It should be noted that β depends on both the window size L and the investment horizon T. When β is large, even a small change in $\hat{X}_t$ would result in a large change in ξ_t^*. Hence, we can view β as a measure of *market timing aggressiveness*. We shall compare the optimal and the IID strategies by an appropriate performance metric discussed as follows.

Given any portfolio strategy $\xi \equiv \{\xi_t\}_{t=0}^T$, we define the *holding period return* (HPR) over the period $[0, T]$ as $\log(W_T^\xi / W_0^\xi)$,

$$\log W_T^\xi - \log W_0^\xi = rT + \int_0^T \left[\xi_t(\mu_t - r) - \xi_t^2 \frac{\sigma^2}{2} \right] \mathrm{d}t + \sigma \int_0^T \xi_t \mathrm{d}B_t, \tag{3.25}$$

where W_t^ξ denotes the level of wealth associated with strategy ξ at time t. Note that while ξ is chosen according to the investor's subject reference model $\tilde{\mathbb{P}}$, the actual dynamics of wealth are determined by the true model $\mathbb{P}$.

Let us then denote the *expected holding period return* by

$$\mathcal{R}_T^\xi \equiv \mathrm{E}[\log W_T^\xi - \log W_0^\xi].$$

This quantity represents the average total return on the portfolio over the period

$[0, T]$ if the investor plays infinite times the investment game ξ. For analytical studies, this metric has two important advantages: first, it is an *objective* performance measure because the expectation is taken with respect to the true probability $\mathbb{P}$ (not $\tilde{\mathbb{P}}$); second, it is independent of any single set of observations. Indeed, looking at expected HPRs is conceptually consistent with performing bootstrapping in many empirical studies. For example, if an econometrician correctly identifies the model specification (3.1)-(3.2) and uses simulation-based approaches to test the profitability of any portfolio strategy ξ, then the simulated HPRs would center around $\mathcal{R}_T^\xi$. Similar results should be obtained by other resampling approaches that reasonably capture the underlying correlations of stock returns. For the rest of this article, we set $W_0 = 1$ such that $\mathcal{R}_T^\xi = \mathrm{E}[\log W_T^\xi]$.

Taking the expectation of (3.25), we obtain the expected HPR formula for any strategy ξ,

$$\mathcal{R}_T^\xi = rT + \mathrm{E}\int_0^T \xi_t(\mu_t - r)\mathrm{d}t - \frac{\sigma^2}{2}\mathrm{E}\int_0^T \xi_t^2 \mathrm{d}t. \tag{3.26}$$

Proposition 3.2. *The gain from expected holding period return from the optimal strategy over the IID strategy is nonnegative, with*

$$\mathcal{R}_T^{\xi^*} - \mathcal{R}_T^{\xi^o} = \frac{1}{2}\frac{\beta^2}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t \geq 0. \tag{3.27}$$

This inequality is strict if β is nonzero.

This proposition states that if β is nonzero, then the optimal strategy ξ^* is better than the IID strategy ξ^o in terms of expected HPR.

Proof. For the IID strategy, substituting (3.23) into the expected HPR formula (3.26), we can show that

$$\mathcal{R}_T^{\xi^o} = rT + \frac{1}{2}\frac{(\overline{\mu} - r)^2}{\sigma^2}T. \tag{3.28}$$

Similarly, for the optimal strategy, using (3.24), we have

$$\begin{aligned}
\mathcal{R}_T^{\xi^*} &= rT + \mathrm{E}\int_0^T \left(\xi_t^o + \frac{\beta \hat{X}_t}{\sigma^2}\right)(\mu_t - r)\mathrm{d}t \\
&\quad - \frac{\sigma^2}{2}\mathrm{E}\int_0^T \left[(\xi_t^o)^2 + \frac{\beta^2}{\sigma^4}\hat{X}_t^2 + 2\xi_t^o\frac{\beta\hat{X}_t}{\sigma^2}\right]\mathrm{d}t \\
&= \left[rT + \mathrm{E}\int_0^T \xi_t^o(\mu_t - r)\mathrm{d}t - \frac{\sigma^2}{2}\mathrm{E}\int_0^T (\xi_t^o)^2\mathrm{d}t\right] \\
&\quad + \frac{\beta}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t\mu_t\mathrm{d}t - \frac{1}{2}\frac{\beta^2}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t^2\mathrm{d}t.
\end{aligned}$$

To obtain the second equality, we apply the property that $\mathrm{E}\int_0^T \hat{X}_t\mathrm{d}t = 0$. By noting the term in the squared brackets is $\mathcal{R}_T^{\xi^o}$ and subtracting

$$\frac{\beta}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t\overline{\mu}\,\mathrm{d}t \equiv 0,$$

we have

$$\begin{aligned}
\mathcal{R}_T^{\xi^*} &= \mathcal{R}_T^{\xi^o} + \frac{\beta}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t(\mu_t - \overline{\mu})\mathrm{d}t - \frac{1}{2}\frac{\beta^2}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t^2\mathrm{d}t \\
&= \mathcal{R}_T^{\xi^o} + \frac{\beta}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t\mathrm{d}t - \frac{1}{2}\frac{\beta^2}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t^2\mathrm{d}t.
\end{aligned} \tag{3.29}$$

Lastly, substituting the definition of β, given by (3.17), after simplification, we obtain (3.27), as desired. ■

We note that this result depends on β being MISE minimizing. If β were simply treated as a parameter to be estimated by data, then we would not be able to simplify (3.29) to obtain (3.27). We also note that the expression (3.27) can be rewritten as

$$\mathcal{R}_T^{\xi^*} - \mathcal{R}_T^{\xi^o} = \frac{1}{\sigma^2}|\mathcal{L}(\beta;T)| \geq 0,$$

which states that the expected gain in HPR is directly proportional to the expected gain in forecast accuracy represented by $|\mathcal{L}(\beta;T)|$.

We now generalize Proposition 3.2 to show that the optimal strategy ξ^* is actually optimal over the class of strategies ξ^b of the form

$$\xi_t^b \equiv \frac{\overline{\mu} + b\hat{X}_t - r}{\sigma^2} = \xi_t^o + \frac{b\hat{X}_t}{\sigma^2}, \tag{3.30}$$

where b is any real number. Note that ξ^b would be the strategy used by a log-utility investor who takes the slope parameter β as b in the linear prediction model (3.9).

Proposition 3.3. *Let ξ^b be the class of strategies as defined by* (3.30). *We have*

$$\mathcal{R}_T^{\xi^*} \geq \mathcal{R}_T^{\xi^b}$$

for any real number b. If $b = \beta$, then we have $\mathcal{R}_T^{\xi^*} = \mathcal{R}_T^{\xi^b}$.

Proof. Similar to (3.29), the expected HPR for ξ^b is given by

$$\mathcal{R}_T^{\xi^b} = \mathcal{R}_T^{\xi^o} + \frac{b}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t\mathrm{d}t - \frac{1}{2}\frac{b^2}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t\mathrm{d}t. \tag{3.31}$$

We note that $\mathcal{R}_T^{\xi^b}$ is a concave quadratic function of b. Differentiating it with respect to b and finding the optimal value of b, denoted by b^*, shows that

$$b^* = \frac{\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t\mathrm{d}t}{\mathrm{E}\int_0^T \hat{X}_t^2\mathrm{d}t},$$

which is precisely the MISE minimizing β. ■

The significance of Proposition 3.3 is that not only does β minimize the MISE, it also maximizes the expected HRP among the class of market timing strategies ξ^b. Hence, β is both statistically and economically optimal to a log-utility investor.

However, our assumption that the investor knows the value of β is very strong. More realistically, the investor would likely need to estimate its value. Let b denote such an estimate and we interpret $e_b \equiv \beta - b$ as the *estimation error*. Under the log-utility assumption, we can imagine that the investor simply takes b as the true value of β and use the portfolio strategy ξ^b. Since it is unlikely that b is precisely β, we are interested to examine whether investors, subject to estimation error, still benefit from market timing.

Proposition 3.4. *Let ξ^b be the class of strategies as defined by* (3.30)*. Assuming, without loss of generality, that* $\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t \mathrm{d}t > 0$ *such that* $\beta > 0$*, then we have*

$$\mathcal{R}_T^{\xi^b} \geq \mathcal{R}_T^{\xi^o},$$

as long as b lies within the interval $0 \leq b \leq 2\beta$*. Or equivalently, the estimation error, $e_b \equiv \beta - b$, lies within the interval*

$$-\beta \leq e_b \leq \beta. \tag{3.32}$$

Proof. By (3.31), we see that $\mathcal{R}_T^{\xi^b} \geq \mathcal{R}_T^{\xi^o}$ if and only if

$$\frac{b}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t \mathrm{d}t - \frac{1}{2}\frac{b^2}{\sigma^2}\mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t \geq 0.$$

But this quadratic inequality holds if and only if

$$0 \le b \le 2\frac{\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t \mathrm{d}t}{\mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t} = 2\beta.$$

This completes the proof. ■

A relevant implication by Proposition 3.4 is that, if stock returns come from an IID model (implying $\beta = 0$ because $\mathrm{E}\int_0^T \hat{X}_t\hat{\mu}_t \mathrm{d}t = 0$), then any market timing strategy ξ^b with $b \neq 0$ is worse than the IID strategy, as implied by the inequality (3.32).

3.3.6 Measuring Window Size L in Days

Before we move on to calibrating our model, we want to find a formula that allows us to measure the window size parameter L in days because moving averages are measured in days in practice. Given any fixed point t in $[0, T]$, we define $t_0 = t$ and $t_{-j} = t_0 - j\Delta t$, $j = 1, 2, \ldots$, for some stepsize Δt. Since T is measured in years, assuming there are 252 trading days a year, we set $\Delta t = \frac{1}{252}$ such that each interval $(t_{-j}, t_{-j+1}]$ represents one trading day. Thus, we can approximate $a_t = \log A_t$, with A_t defined in (3.4), by

$$a_t \approx L \sum_{j=1}^{\infty} e^{-L(j-1)\Delta t} p_{t-j}\Delta t.$$

Comparing this expression with the N-day (geometric) exponentially weighted moving average in discrete time,

$$\sum_{j=1}^{\infty} \left(\frac{2}{N+1}\right)\left(1 - \frac{2}{N+1}\right)^{j-1} p_{t-j},$$

we can relate L to N by setting

$$Le^{-L\Delta t(j-1)}\Delta t = \left(\frac{2}{N+1}\right)\left(1-\frac{2}{N+1}\right)^{j-1}.$$

Using the approximation $e^{-L\Delta t(j-1)} \approx (1-L\Delta t)^{j-1}$, we have

$$L \approx \frac{2}{N+1}\frac{1}{\Delta t} = \frac{504}{N+1}.$$

This formula is handy. For example, if we consider 50-day MA indicators then we set $L = 504/51 \approx 9.882$. Similarly, for 100-day MA indicators, we set $L = 504/101 \approx 4.990$. We see that the longer the moving average in discrete time, the smaller the window size in continuous time.

3.4 Data and Model Calibration

For simplicity, let us assume that the percentage drift is linearly related to a zero-mean state variable δ_t such that we have

$$\mu_t = \overline{\mu} + \theta\delta_t,$$

where θ is a constant parameter. This assumption implies that the stochastic process determining the cum-dividend stock prices, given by (3.1)-(3.2), can be rewritten as follows,

$$\frac{\mathrm{d}P_t}{P_t} = (\overline{\mu} + \theta\delta_t)\mathrm{d}t + \sigma\mathrm{d}B_t, \tag{3.33}$$

$$\mathrm{d}\delta_t = -\lambda\delta_t\mathrm{d}t + \omega\mathrm{d}Z_t, \tag{3.34}$$

where $\omega \equiv \eta/\theta$ and $\mathrm{d}B_t\mathrm{d}Z_t = \rho\,\mathrm{d}t$.

We estimate this joint stochastic process by using a discrete approximation to the continuous processes, and using monthly data for the period January 1980 to December 2014. The stock return is taken as the rate of return on S&P 500 index and the state variable is taken as detrended log dividend yield[1]. The system of equations (3.33)-(3.34) is estimated by linear seemingly unrelated regression (SUR). We present the estimation results in Table 3.1.

Table 3.1. Results of Parameter Estimation

This table lists parameter notations, estimated values, and relevant statistics. The parameters are estimated from monthly data of stock returns and dividend yield for the period January 1980 to December 2014. The stock return, $\mathrm{d}P_t/P_t$, is taken as the rate of return on S&P 500 index and the state variable, δ_t, is taken as detrended log dividend yield.

Parameter	Estimated value	Standard error	t-value	95% CI lower bound	95% CI upper bound
$\overline{\mu}$	0.217	0.044	4.888	0.130	0.303
θ	1.701	0.428	3.976	0.862	2.539
σ	0.198	0.010	19.833	0.179	0.218
λ	1.332	0.360	3.701	0.626	2.037
ω	0.167	0.010	16.064	0.146	0.187
η	0.284	0.109	2.593	0.069	0.498
ρ	-0.080	0.068	-1.177	-0.214	0.054

Notes: The standard errors of the estimated $(\sigma, \omega, \rho, \eta)$ are bootstrapped standard errors using 15,000 bootstrap samples. The rest are analytical (asymptotic) standard errors.

The coefficient of determination (or R^2) for the discretized regression (3.33) is 2.740%, and 3.196% for (3.34). We also note that ideally one could improve the estimation by using nonlinear SUR but we believe that our estimation gives us reasonable parameter estimates for numerical illustrations.

[1] We detrend log dividend yield by a Hodrick-Prescott filter to ensure stationarity. The smoothing parameter is set to be 129, 600 for monthly data, as recommended by Ravn and Uhlig (2002).

3.5 Numerical Results

3.5.1 MISE Minimizing β

We recall that the definition of β is given by (3.17) and its computation is discussed in Section 3.3.3. For example, consider the case of 50-day moving averages with $L \approx 9.882$, after substituting the estimated parameters into the algebraic solutions of v_t (not reported in this article), the solutions of v_t can be written as follows:

$$\begin{aligned}
\mathrm{E}[p_t^2] &= 0.0685t + 0.0477e^{-1.33t} - 0.0129e^{-2.67t} + 0.0388t^2 - 0.0348, \\
\mathrm{E}[p_t a_t] &= 0.00686t + 0.00486e^{-1.33t} - 0.00132e^{-2.67t} - 0.00312e^{-97.7t} \\
&\quad + 0.00316e^{-99.0t} + 0.00392t^2 - 0.00357, \\
\mathrm{E}[p_t \mu_t] &= 0.0426t - 0.0318e^{-1.33t} + 0.0172e^{-2.67t} + 0.0146, \\
\mathrm{E}[a_t^2] &= 0.00069t + 0.000495e^{-1.33t} - 0.000136e^{-2.67t} - 0.000632e^{-97.7t} \\
&\quad + 0.000386e^{-195t} + 0.000648e^{-99.0t} + 0.000397t^2 - 0.000365, \\
\mathrm{E}[a_t \mu_t] &= 0.000483t - 0.00322e^{-1.33t} + 0.00177e^{-2.67t} - 0.00421e^{-99.0t} \\
&\quad + 0.00135, \\
\mathrm{E}[\mu_t^2] &= 0.0698 - 0.023e^{-2.67t},
\end{aligned}$$

where we round all numbers to three significant digits. We note that the solutions of v_t corresponding to other moving-average lag days share the same functional forms as above. The solutions of v_t for all moving averages (10, 50, 100, and 200 lag days) considered in this article are summarized in Appendix 3.D.

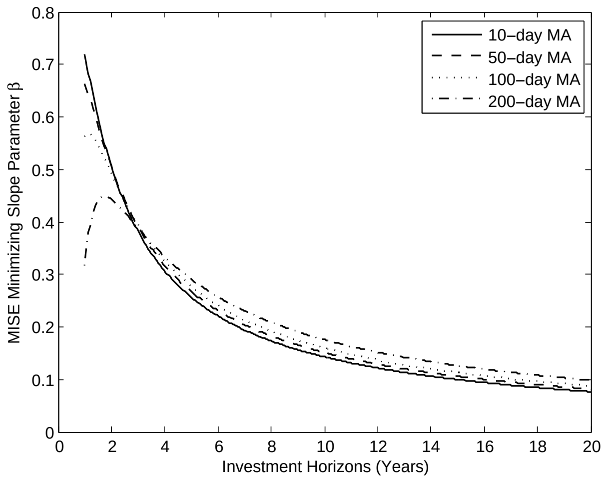

Figure 3.1. Mean Integrated Squared Error (MISE) Minimizing Slope Parameter β. This figure plots the "best" slope parameter β of the investor's return prediction model as a function of investment horizon ranging from 1 to 20 years. We compare β for four moving average indicators ranging from 10 to 200 days.

Figure 3.1 shows the MISE minimizing β for four moving averages as a function of investment horizon, ranging from one to twenty years. Two aspects of this figure are noteworthy. First, β is nonzero, this means that the MA indicators have predictive power on stock returns in the sense that using the best linear predictor $\overline{\mu} + \beta \hat{X}_t$ improves expected forecast accuracy over the long-run mean $\overline{\mu}$ alone. In particular, our empirical result that $\beta > 0$ is consistent with the conventional practice that technical investors increase allocation to stocks when the current stock price is sufficiently higher than the moving averages. Second, β eventually declines in horizons, although such declines can be nonmonotonic. Since β also measures market timing aggressiveness, our results imply that, in general, the optimal allocations to stocks of shorter-horizon investors tend to be sensitive to even small changes in MA indicators, compared to longer-horizon investors (say, five years or more). Also, recall the inequality (3.32) in Proposition 3.4 that we require the estimation error to satisfy $-\beta \leq e_b \leq \beta$ for a gain in expected HRP. Hence, a larger β implies a wider interval. This means that, in general, the portfolio profitability of shorter-horizon investors is more robust to estimation errors.

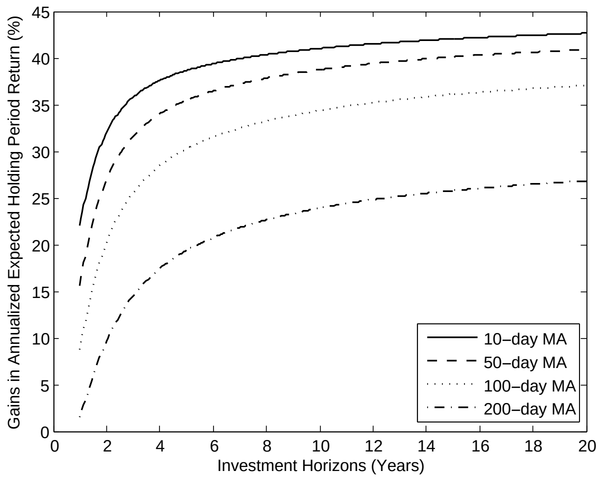

Figure 3.2. Gains in Annualized Expected Holding Period Return (HPR). This figure plots the gains in annualized expected HPRs for the optimal strategy ξ^* over the IID strategy ξ^o, given by $(\mathcal{R}_T^{\xi^*} - \mathcal{R}_T^{\xi^o})/T$, as a function of investment horizon T, ranging from 1 to 20 years. We compare four moving average indicators ranging from 10 to 200 days.

3.5.2 Gains in Annualized Expected Holding Period Return

Figure 3.2 plots the *gain in annualized expected HPR* for the optimal strategy over the IID strategy, defined as $(\mathcal{R}_T^{\xi^*} - \mathcal{R}_T^{\xi^o})/T$, as a function of horizon, ranging from one to twenty years. Note that the annualized expected HPR for the IID strategy is a constant, $r + \frac{1}{2}(\overline{\mu} - r)^2/\sigma^2$, as implied by (3.28). Two aspects of this figure are noteworthy. First, the gain is economically significant, meaning that one can expect the optimal technical strategies to substantially outperform the IID strategy. However, we emphasize that this result depends on the assumption that the investor knows the true value of β. When β is incorrectly identified, such expected gain can substantially reduce, or even become negative. Second, expected gain increases with horizons and eventually converges. This finding shows that the optimal strategies can have long-run profitability even though long-term investors use a rather mild market timing approach. Although this finding seems encouraging, we recall that β declines with horizons quickly and so investors not knowing the true β are more likely to suffer from an estimation error large enough to erode any expected gain. Third, we see that shorter moving averages have higher annualized expected gains. This result is likely due to the model assumption that the true proportional drift of stock prices is Markovian and so shorter moving averages better capture the local drift. However, observe also that the expected gains are similar across the 10 and 50-day moving averages. Thus, our model implies that there exists a range of moving averages that are of similar profitability in terms of expected HPR. This finding seems to be consistent with the fact that investors do not systematically favor a particular moving average in practice.

3.5.3 Comparative Statics

Using 50-day MA indicators as an example, we provide comparative statics to examine how some key model parameters of the actual stock return model affect the gains in annualized expected HPR associated with the optimal technical strategy. The results for other moving averages are qualitatively similar so we omit them for brevity.

First, we consider the role of the predictive power of the state variable δ_t, represented by θ, on gains in annualized expected HPR. Recall from (3.34) that the volatility of the proportional drift μ_t, represented by η, relates to θ by $\eta = \theta\omega$. Panel A of Figure 3.3 compares the gains in annualized expected HPR evaluated at the mean estimate, 95% confidence interval lower and upper bounds of θ, as reported in Table 3.1, while fixing ω to its mean estimate. Our results show that such expected gain is increasing in the predictive power of the state variable. This finding confirms the intuition that the optimal technical strategies can improve expected HPR due to stock return predictability.

Next, we investigate the role of the volatility of stock returns, represented by σ, on gains in annualized expected HPR. Panel B of Figure 3.3 compares the gains in annualized expected HPR evaluated at the mean estimate, 95% confidence interval lower and upper bounds of σ, as reported in Table 3.1. Our results show that such expected gain is decreasing in the volatility of stock returns. The intuition is that the correlation between the MA indicators and the drift of stock returns is weaken by more noise as implied by a higher volatility.

Finally, we investigate the role of the correlation between the Brownian shocks of stock returns and the state variable, represented by ρ, on gains in annualized expected HPR. Panel C of Figure 3.3 compares the gains in annualized expected HPR evaluated at the mean estimate, 95% confidence interval lower and upper bounds of ρ, as reported in Table 3.1. Our results show that such expected gain is increasing in the correlation between the shocks of stock returns and the state variable. The intuition is that when such a correlation is high, shocks to the stock returns contain more information about the shocks to the drift. Hence, it becomes easier to detect the local movement of the unobservable drift.

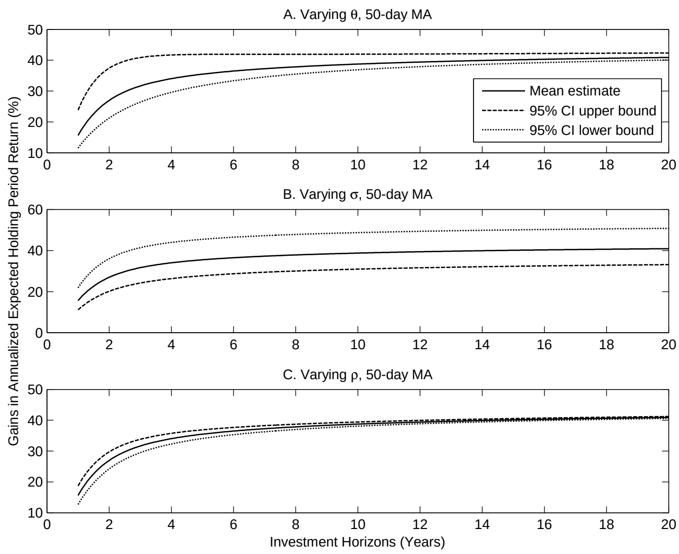

Figure 3.3. Comparative Statics for Gains in Annualized Expected Holding Period Return (HPR). We consider 50-day moving average indicators as an example. This figure illustrates the sensitivity of the gain in annualized expected HPR for the optimal strategy over the IID strategy to the model parameters θ, σ, and ρ, which represent the predictive power of the state variable δ_t, the volatility of stock returns, and the correlation between the Brownian shocks of stock returns and the state variable.

3.6 Conclusion

In practice, it is difficult to identity the state variable determining the drift of stock returns. Indeed, most predictive variables studied in the literature cannot be observed at a desired frequency for real-time trading. It is also difficult to identify the stochastic process determining the state variable. In this article, we present a portfolio choice model to theoretically illustrate that moving averages (MAs) can be useful for investment when stock returns are correlated. Our modelling strategy is as follows. First, we propose a measure of predictive power in the context of linear prediction models in continuous time, defined over a forecast horizon and a loss function. Next, assuming stock returns follow a state-space model (which implies serial correlations), we show that MAs have predictive power for returns in the sense that the associated best linear predictor improves expected forecast accuracy over the long-run mean of returns alone. Then, we form MA-based optimal portfolio strategies and show their expected profitability. After that, we calibrate our model with S&P 500 index and dividend yield data to illustrate the economic significance of our results. Finally, we provide comparative statics to examine how some key model parameters of the state-space model for stock returns affect the profitability of our strategies.

3.7 Appendix

Appendix 3.A. Conditional Log-Likelihood Function

Consider approximating (3.9) with the Euler difference scheme

$$R_{t_{j+1}} \equiv \frac{P_{t_{j+1}} - P_{t_j}}{P_{t_j}} = (\overline{\mu} + \beta \hat{X}_{t_j})\Delta t + \sigma\sqrt{\Delta t}\,\epsilon_{t_j},$$

for $j = 0, 1, \ldots, n-1$ where $\Delta t \equiv T/n$ and ϵ_{t_j} are IID standard Gaussian random variables. Given the data set $\mathcal{F}_T \equiv \{(R_{t_{j+1}}, \hat{X}_{t_j})\}_{j=0}^{n-1}$, we can write down the conditional likelihood function

$$\mathrm{L}(\beta|\mathcal{F}_T) \equiv \prod_{j=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left\{-\frac{(R_{t_{j+1}} - (\overline{\mu} + \beta\hat{X}_{t_j})\Delta t)^2}{2\sigma^2\Delta t}\right\}.$$

Thus, the log-likelihood function is

$$\begin{aligned}
\log \mathrm{L}(\beta|\mathcal{F}_T) &\propto \sum_{j=0}^{n-1} -\frac{R_{t_{j+1}}^2 + (\overline{\mu} + \beta\hat{X}_{t_j})^2(\Delta t)^2 - 2R_{t_{j+1}}(\overline{\mu} + \beta\hat{X}_{t_j})\Delta t}{2\sigma^2\Delta t} \\
&\propto -\frac{1}{2}\sum_{j=0}^{n-1}(\overline{\mu} + \beta\hat{X}_{t_j})^2\Delta t + \sum_{j=0}^{n-1}(\overline{\mu} + \beta\hat{X}_{t_j})R_{t_{j+1}} \\
&\propto -\frac{1}{2}\beta^2\sum_{j=0}^{n-1}\hat{X}_{t_j}^2\Delta t - \overline{\mu}\beta\sum_{j=0}^{n-1}\hat{X}_{t_j}\Delta t + \beta\sum_{j=0}^{n-1}\left(\hat{X}_{t_j}\tfrac{P_{t_{j+1}} - P_{t_j}}{P_{t_j}}\right),
\end{aligned}$$

where the notation "$\propto$" means "is proportional to".

Taking the limit as $n \to \infty$, we have

$$\log \mathrm{L}(\beta|\mathcal{F}_T) \propto -\frac{1}{2}\beta^2\int_0^T \hat{X}_t^2 \mathrm{d}t - \overline{\mu}\beta\int_0^T \hat{X}_t \mathrm{d}t + \beta\int_0^T \hat{X}_t\frac{\mathrm{d}P_t}{P_t}.$$

Taking expectation, we obtain

$$
\begin{aligned}
\mathrm{E}[\log \mathrm{L}(\beta|\mathcal{F}_T)] &\propto -\frac{1}{2}\beta^2 \mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t - \overline{\mu}\beta \mathrm{E}\int_0^T \hat{X}_t \mathrm{d}t + \beta \mathrm{E}\int_0^T \hat{X}_t \frac{\mathrm{d}P_t}{P_t} \\
&= -\frac{1}{2}\beta^2 \mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t - \overline{\mu}\beta \mathrm{E}\int_0^T \hat{X}_t \mathrm{d}t + \beta \mathrm{E}\int_0^T \hat{X}_t \mu_t \mathrm{d}t \\
&= -\frac{1}{2}\beta^2 \mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t + \beta \mathrm{E}\int_0^T \hat{X}_t (\mu_t - \overline{\mu}) \mathrm{d}t,
\end{aligned}
$$

which is precisely the negative of $\mathcal{L}(\beta; T)$, given by (3.16), as claimed. Thus, we see that choosing β by minimizing the MISE is equivalent to that of maximizing expected log-likelihood and this result does not require $X_0 = \overline{X}$.

Appendix 3.B. MISE Minimizing Parameters (α, β)

Consider the linear prediction model (3.8), reproduced here for convenience,

$$
\frac{\mathrm{d}P_t}{P_t} = (\alpha + \beta X_t)\mathrm{d}t + \sigma \mathrm{d}\tilde{B}_t.
$$

For any fixed horizon T, we want to show that the MISE minimizing values of (α, β) are

$$
\alpha = \overline{\mu} - \beta \overline{X} \quad \text{and} \quad \beta = \frac{\mathrm{E}\int_0^T \hat{X}_t \hat{\mu}_t \mathrm{d}t}{\mathrm{E}\int_0^T \hat{X}_t^2 \mathrm{d}t},
$$

if we assume X_0 starts at its long-run mean, *i.e.*, $X_0 = \overline{X}$, such that

$$
\mathrm{E}\int_0^T \hat{X}_t \mathrm{d}t = \mathrm{E}\int_0^T (X_t - \overline{X})\mathrm{d}t = 0.
$$

Following the definition of MISE in (3.11), we want to find (α, β) to minimize the quantity

$$
\mathrm{MISE} = \mathrm{E}\int_0^T \left[\frac{P_{t+h} - P_t}{P_t} - (\alpha + \beta X_t)h - o(h)\right]^2 \mathrm{d}t.
$$

Following a procedure similar to that of Section 3.3.2, this minimization problem can be reduced to

$$\min_{(\alpha,\beta)} \mathcal{L}(\alpha,\beta;T) \equiv \frac{1}{2}\mathrm{E}\int_0^T (\alpha+\beta X_t)^2 \mathrm{d}t - \mathrm{E}\int_0^T (\alpha+\beta X_t)\frac{\mathrm{d}P_t}{P_t}.$$

Thus, the first-order conditions are

$$\mathrm{E}\int_0^T (\alpha+\beta X_t)\mathrm{d}t - \mathrm{E}\int_0^T \frac{\mathrm{d}P_t}{P_t} = 0, \tag{3.35}$$

and

$$\mathrm{E}\int_0^T (\alpha+\beta X_t)X_t\mathrm{d}t - \mathrm{E}\int_0^T X_t\frac{\mathrm{d}P_t}{P_t} = 0. \tag{3.36}$$

Simplifying (3.35), we have

$$\alpha T + \beta\mathrm{E}\int_0^T X_t\mathrm{d}t - \overline{\mu}T = 0,$$

and rearranging gives

$$\alpha = \overline{\mu} - \beta\frac{1}{T}\mathrm{E}\int_0^T X_t\mathrm{d}t = \overline{\mu} - \beta\overline{X}.$$

Substituting this into (3.36), we have

$$\begin{aligned}\mathrm{E}\int_0^T [\overline{\mu}+\beta(X_t-\overline{X})]X_t\mathrm{d}t - \mathrm{E}\int_0^T X_t\mu_t\mathrm{d}t = 0\\ \beta\mathrm{E}\int_0^T \hat{X}_t X_t\mathrm{d}t - \mathrm{E}\int_0^T X_t\hat{\mu}_t\mathrm{d}t = 0,\end{aligned}$$

where we use the definitions $\hat{X}_t \equiv X_t - \overline{X}$ and $\hat{\mu}_t \equiv \mu_t - \overline{\mu}$. Finally, adding

$$\mathrm{E}\int_0^T \overline{X}\hat{\mu}_t \mathrm{d}t - \beta\mathrm{E}\int_0^T \hat{X}_t\overline{X}\mathrm{d}t \equiv 0,$$

we have

$$\beta\mathrm{E}\int_0^T \hat{X}_t(X_t - \overline{X})\mathrm{d}t - \mathrm{E}\int_0^T (X_t - \overline{X})\hat{\mu}_t \mathrm{d}t = 0.$$

This gives the desired solution for β.

Appendix 3.C. Definitions of A_0 and A_1 in Section 3.3.4

We present the definitions of A_0 and A_1 in the differential equation (3.22)

$$\frac{\mathrm{d}\phi_t}{\mathrm{d}t} = A_0 + A_1\phi_t$$

as follows. We have

$$A_0 = (-\sigma^2/2, 0, \lambda\overline{\mu}, \sigma^2, 0, \eta\rho\sigma, 0, 0, \eta^2)^\top,$$

and

$$A_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ L & -L^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sigma^2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -\sigma^2/2 & 0 & L & -L^2 & 0 & 0 & 1 & 0 \\ \lambda\overline{\mu} & 0 & -\sigma^2/2 & 0 & 0 & -\lambda & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2L & 0 & -2L^2 & 0 & 0 \\ 0 & \lambda\overline{\mu} & 0 & 0 & 0 & L & 0 & -(L^2+\lambda) & 0 \\ 0 & 0 & 2\lambda\overline{\mu} & 0 & 0 & 0 & 0 & 0 & -2\lambda \end{bmatrix}.$$

Appendix 3.D. Solutions to the System of ODEs $\mathrm{d}v_t/\mathrm{d}t$ in Section 3.3.3

Based on the estimated parameters in Table 3.1, we compute the solutions to $\mathrm{d}v_t/\mathrm{d}t$, equation (20), for the cases of 10, 50, 100, and 200-day moving averages. The corresponding values of the window size parameter L are 45.818, 9.882, 4.990, and 2.507.

For 10-day MA ($L \approx 45.818$),

$$
\begin{aligned}
\mathrm{E}[p_t^2] &= 0.0685t + 0.0477e^{-1.33t} - 0.0129e^{-2.67t} + 0.0388t^2 - 0.0348, \\
\mathrm{E}[p_t a_t] &= 0.00149t - 0.000697e^{-2100.0t} + 0.000697e^{-2100.0t} + 0.00104e^{-1.33t} \\
&\qquad - 0.000282e^{-2.67t} + 0.000846t^2 - 0.000760, \\
\mathrm{E}[p_t \mu_t] &= 0.0426t - 0.0318e^{-1.33t} + 0.0172e^{-2.67t} + 0.0146, \\
\mathrm{E}[a_t^2] &= 0.0000326t - 0.0000304e^{-2100.0t} + 0.0000184e^{-4200.0t} \\
&\qquad + 0.0000304e^{-2100.0t} + 0.0000227e^{-1.33t} - 0.00000616e^{-2.67t} \\
&\qquad\quad + 0.0000185t^2 - 0.0000166, \\
\mathrm{E}[a_t \mu_t] &= 0.000957t - 0.000929e^{-2100t} - 0.000694e^{-1.33t} + 0.000376e^{-2.67t} \\
&\qquad + 0.000317, \\
\mathrm{E}[\mu_t^2] &= 0.0698 - 0.023e^{-2.67t}.
\end{aligned}
$$

For 50-day MA ($L \approx 9.882$),

$$\begin{aligned}
\mathrm{E}[p_t^2] &= 0.0685t + 0.0477e^{-1.33t} - 0.0129e^{-2.67t} + 0.0388t^2 - 0.0348, \\
\mathrm{E}[p_t a_t] &= 0.00686t + 0.00486e^{-1.33t} - 0.00132e^{-2.67t} - 0.00312e^{-97.7t} \\
&\quad + 0.00316e^{-99.0t} + 0.00392t^2 - 0.00357, \\
\mathrm{E}[p_t \mu_t] &= 0.0426t - 0.0318e^{-1.33t} + 0.0172e^{-2.67t} + 0.0146, \\
\mathrm{E}[a_t^2] &= 0.00069t + 0.000495e^{-1.33t} - 0.000136e^{-2.67t} - 0.000632e^{-97.7t} \\
&\quad + 0.000386e^{-195t} + 0.000648e^{-99.0t} + 0.000397t^2 - 0.000365, \\
\mathrm{E}[a_t \mu_t] &= 0.000483t - 0.00322e^{-1.33t} + 0.00177e^{-2.67t} - 0.00421e^{-99.0t} \\
&\quad + 0.00135, \\
\mathrm{E}[\mu_t^2] &= 0.0698 - 0.023e^{-2.67t}.
\end{aligned}$$

For 100-day MA ($L \approx 4.990$),

$$\begin{aligned}
\mathrm{E}[p_t^2] &= 0.0685t + 0.0477e^{-1.33t} - 0.0129e^{-2.67t} + 0.0388t^2 - 0.0348, \\
\mathrm{E}[p_t a_t] &= 0.0134t + 0.00983e^{-1.33t} - 0.00273e^{-2.67t} - 0.00552e^{-24.9t} \\
&\quad + 0.00582e^{-26.2t} + 0.00777t^2 - 0.0074, \\
\mathrm{E}[p_t \mu_t] &= 0.0426t - 0.0318e^{-1.33t} + 0.0172e^{-2.67t} + 0.0146, \\
\mathrm{E}[a_t^2] &= 0.00261t + 0.00202e^{-1.33t} - 0.000579e^{-2.67t} - 0.00221e^{-24.9t} \\
&\quad + 0.00139e^{-49.8t} + 0.00247e^{-26.2t} + 0.00156t^2 - 0.00153, \\
\mathrm{E}[a_t \mu_t] &= 0.0103t - 0.00637e^{-1.33t} + 0.00365e^{-2.67t} - 0.00776e^{-26.2t} \\
&\quad + 0.00195, \\
\mathrm{E}[\mu_t^2] &= 0.0698 - 0.023e^{-2.67t}.
\end{aligned}$$

For 200-day MA ($L \approx 2.507$),

$$\begin{aligned}
\mathrm{E}[p_t^2] &= 0.0685t + 0.0477e^{-1.33t} - 0.0129e^{-2.67t} + 0.0388t^2 - 0.0348, \\
\mathrm{E}[p_t a_t] &= 0.0252t + 0.0216e^{-1.33t} - 0.00654e^{-2.67t} - 0.00628e^{-6.29t} \\
&\quad + 0.00903e^{-7.62t} + 0.0155t^2 - 0.0178, \\
\mathrm{E}[p_t \mu_t] &= 0.0426t - 0.0318e^{-1.33t} + 0.0172e^{-2.67t} + 0.0146, \\
\mathrm{E}[a_t^2] &= 0.00906t + 0.00963e^{-1.33t} - 0.00331e^{-2.67t} - 0.00501e^{-6.29t} \\
&\quad + 0.00353e^{-12.6t} + 0.00914e^{-7.62t} + 0.00616t^2 - 0.00782, \\
\mathrm{E}[a_t \mu_t] &= 0.0215t - 0.0127e^{-1.33t} + 0.00871e^{-2.67t} - 0.012e^{-7.62t} \\
&\quad - 0.000995, \\
\mathrm{E}[\mu_t^2] &= 0.0698 - 0.023e^{-2.67t}.
\end{aligned}$$

We note that for each case, the solutions of $\mathrm{E}[p_t^2]$, $\mathrm{E}[p_t \mu_t]$, and $\mathrm{E}[\mu_t^2]$ are the same because both p_t and μ_t, being exogenous processes, are independent of the choice of L.

3.8 References

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? Review of Financial Studies, 20(3), 651-707.

Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. Journal of Financial Economics, 51(2), 245-271.

Bajgrowicz, P., & Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics, 106(3), 473-491.

Barberis, N. (2000). Investing for the long run when returns are predictable. Journal of Finance, 55(1), 225-264.

Boudoukh, J., Michaely, R., Richardson, M., & Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. Journal of Finance, 62(2), 877-915.

Brennan, M. J., Schwartz, E. S., & Lagnado, R. (1997). Strategic asset allocation. Journal of Economic Dynamics and Control, 21(8), 1377-1403.

Brock, W. A., & Malliaris, A. G. (1992). Differential equations, stability and chaos in dynamic economics, Advanced Textbooks in Economics (Book 27), North Holland.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

Campbell, J. Y., & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies, 1(3), 195-228.

Diebold, F. X., & Kilian, L. (2001). Measuring predictability: Theory and macroeconomic applications. Journal of Applied Econometrics, 16(6), 657-669.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis, 48(05), 1433-1461.

Kandel, S., & Stambaugh, R. F. (1995). On the predictability of stock returns: An asset-allocation perspective. Journal of Finance, 51(2), 385-424.

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. Journal of Financial Economics, 44(2), 169-203.

Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies, 3(3), 431-467.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. Journal of Finance, 55(4), 1705-1770.

Neftci, S. N. (1991). Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory: A study of "technical analysis". Journal of Business, 64(4), 549-571.

Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. Review of Economics and Statistics, 84(2), 371-376.

Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance, 54(4), 1647-1691.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126.

Xia, Y. (2001). Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. Journal of Finance, 56(1), 205-246.

Zhou, G., Zhu, Y., & Qiang, S. (2012). Asset allocation: Can technical analysis add value? International Journal of Portfolio Analysis and Management, 1(1), 43-58.

Zhu, Y., & Zhou, G. (2009). Technical analysis: An asset allocation perspective on the use of moving averages. Journal of Financial Economics, 92(3), 519-544.

Chapter 4

A Dynamic Model of Asset Prices and Technical Forecasting with Moving Averages

This article is a theoretical study of the predictive power of moving averages of past stock prices on future prices as an equilibrium phenomenon. Adapting Lucas' (1978) one-tree representative investor model, we derive and characterize equilibrium prices under different assumptions on the investor's model for forecasting stock prices. We provide proof of equilibrium stability to show whether the investor can learn these equilibria (under different assumptions) by standard least-squares learning rules.

4.1 Introduction

While several empirical studies find evidence for the predictive power of moving averages on stock returns, see, *e.g.* Brock, Lakonishok, and LeBaron (1992), Han, Yang, and Zhou (2013), and Neely *et al.* (2014), to the best of our knowledge, Zhou and Zhu (2013) provide the only theoretical equilibrium model that explicitly demonstrates the existence of such predictive power as an equilibrium phenomenon. However, Zhou and Zhu rely on the rational expectations (RE) assumption, which may not be innocuous for reasons explained as follows (see, *e.g.* Evans and Honkapohja (2001) for a thorough discussion).

One major difficulty is that the self-referential feature of an asset pricing model under RE makes multiple equilibria possible, and thus it is important to examine how a particular rational expectations equilibrium (REE) may be arrived at and whether all equilibria are equally worth studying. A second major problem is that it neglects the notion that investors may only learn adaptively to form RE. Indeed, the study of adaptive learning is often used as a selection criterion to reduce the number of attainable (also called learnable, E-stable, or stable) REEs (see, *e.g.* Timmermann, 1994, 1996; Barsky and DeLong, 1993; Branch and Evans, 2010). A third major problem is that it neglects the notion that investors may have misspecified forecasting models but within the context of their subjective model they are unable to detect their misspecification (see, *e.g.* Sargent, 1999; Williams, 2004). Since Zhou and Zhu also assume perfect RE, it is of theoretical interest to justify the value of moving averages by another equilibrium model that takes into account of these issues concerning RE, which this article attempts to provide.

In this article, we replace RE with adaptive learning and provide a theoretical examination of whether moving averages can be useful to forecast future stock prices, as an equilibrium phenomenon, and whether investors who update their expectations by the least-squares approach can learn the equilibrium. Our modelling approach is an adaption of Lucas' (1978) one-tree model populated by a representative investor with a power utility function. We assume that the logarithm of dividends follow a stationary autoregressive process. The main objective of this article is to characterize equilibrium stock prices when we make different assumptions on the investor's forecasting model to predict future prices. From a modelling perspective, our model differs substantially from that of Zhou and Zhu, which is based on Wang (1993).

The model has several interesting implications. First, as is well-known in the literature, if the investor correctly specifies his forecasting model and perceives future prices as depending on dividends only, then there exists a unique learnable REE. Second, if the investor overparameterizes his forecasting model and perceives future prices as depending on both dividends and moving averages of past prices, then there exist two possible equilibria, one of which is precisely the REE identified above while the other is a self-confirming equilibrium implying that stock prices can be predicted by moving averages. However, we find that the second equilibrium is not learnable, meaning that the investor will eventually learn that moving averages are redundant in this scenario.

Third, we assume stock prices are generated according to the REE but consider a measure-zero technical investor (interpreted as a minority with no impact on the economic system) who misspecifies his forecasting model and perceives stock prices as depending on moving averages only. We show that his misspecified model is actually statistically useful in the sense that the optimal slope parameter is nonzero while the forecast errors are orthogonal to the predictor, namely the moving averages. Finally, we consider the scenario when the representative investor is a technical trader who forecasts stock prices as described above. We show that it is possible to have an equilibrium, known as a restricted perceptions equilibrium[1] (RPE), in which the investor cannot detect his misspecification but his expectations are otherwise optimal within a limited class of forecasting models. We find that a sufficient condition for this RPE to be learnable is that when the investor uses a sufficiently short lookback period to compute the moving averages.

Our model contributes to the literature on providing a theoretical basis for the widespread use of technical analysis (see, *e.g.* Treynor and Ferguson, 1985; Brown and Jennings, 1989; Grundy and McNichols, 1989; Blume, Easley, and O'Hara, 1994; Zhou and Zhu, 2013). Our study also relates to the literature on asset pricing models with naive learning, including the least-squares learning approach, for example, Hansen and Sargent (1982), Marcet and Sargent (1989a, 1989b), Barsky and DeLong (1993), Timmermann (1994, 1996), and Branch and Evans (2010). Besides the least-squares learning approach, the Bayesian learning approach is also commonly considered in asset pricing models. For example, Brennan and Xia (2001), Dothan and Feldman (1986), Feldman (2002, 2003, 2007), Lewellen and Shanken (2002), Pástor, Veronesi (2003), and Veronesi (2000). However, while showing the existence (and/or uniqueness) of the equilibrium, asset pricing model builders often pay relatively less attention to prove equilibrium stability, which we demonstrate in our model that existence need not imply that investors can actually learn the equilibrium by standard adaptive learning rules. Indeed, relative to the finance literature, the notion of supplementing an equilibrium study with an investigation of its stability under learning is more common in the macroeconomic literature (see, *e.g.* Evans, 1985; Howitt, 1992; Bullard and Mitra, 2002;

[1] We follow the terminology by Evans and Honkapohja (2001).

Evans and Honkapohja, 2003). Our study, therefore, also serves as an example of explicitly examining the equilibrium stability of an asset pricing model.

Formulating a general equilibrium model is a natural next step to gain more insights about the market pricing mechanism if *everyone*, not just an individual, adapts some technical-based forecasting strategy. This gives us an important justification why returns are serially-correlated in the first place which has been taken as an assumption or "empirical fact" so far.

The asset pricing model in this chapter (and the next related chapter) is formulated in discrete time, not continuous time as adopted in the last two chapters on portfolio choice models. This change is necessary. The major reason is due to the application of some critical mathematical theorems required to study some "learning" problems of interest. Specifically, the E-stability techniques by Evans and Honkapohja (2001) are based on discrete time and so we cannot be sure similar techniques also hold in continuous time. Another reason for this change is that we will use actual dividend data to formally estimate and test a similar asset pricing model in the next chapter. The nature of formal econometric assessment is very different from the model calibration in Chapter 3 in which dividend data are only used to estimate parameter values but do not directly enter into any model restrictions to be tested. Since our dividend data are observed monthly, and thus they are too "sparse" for the continuous-time setting to be relevant. Nonetheless, we would like to note that both discrete time and continuous time are equally useful depending on the problem in hand.

4.2 The Model

4.2.1 The Basic Setting

Consider a standard pure-exchange economy (Lucas, 1978) with a single consumer, interpreted as a *representative investor* "stand in" for a large number of identical investors, and a single risky asset. Shares held during period $t-1$, θ_{t-1}, yield a dividend payment D_t at time t; time-t share prices are P_t. The investor wants to maximize expected lifetime utility by financing consumption C_t from an exogenous stochastic dividend stream and proceeds from sales of shares. The utility maximization problem of the representative investor is given by

$$\max_{\{C_t,\theta_t\}_{t=0}^{\infty}} \mathrm{E} \sum_{t=0}^{\infty} \beta^t U(C_t),$$

subject to the budget constraint

$$W_t \equiv (P_t + D_t)\theta_{t-1} = C_t + P_t\theta_t,$$

where $\beta \in (0,1)$ is a discount factor, $U(\cdot)$ is a current period utility function, and E is an expectations operator.

The *value function* associated with this optimization problem is

$$V(W_t) = \max_{\{\theta_t\}}\{U(W_t - P_t\theta_t) + \beta \mathrm{E}_t[V(W_{t+1})]\},$$

where the notation E_t indicates the expectation is taken conditional on the investor's time-t information set.

The first-order condition with respect to θ_t is

$$\frac{\partial V(W_t)}{\partial \theta_t} = U'(C_t)(-P_t) + \beta \mathrm{E}_t \left[V'(W_{t+1})(P_{t+1} + D_{t+1})\right] = 0,$$

where U' and V' denote the first derivatives of U and V. By the envelope theorem, we have

$$V'(W_{t+1}) = U'(C_{t+1}).$$

In the equilibrium, we have $\theta_t = \theta_{t-1}$ such that $C_t = D_t$. This gives us the well-known *equilibrium pricing equation*

$$P_t = \beta \mathrm{E}_t \left[\frac{U'(D_{t+1})}{U'(D_t)}(D_{t+1} + P_{t+1})\right].$$

Assuming the investor has constant relative risk aversion utility function,

$$U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma},$$

for a constant γ with $\gamma > 0$ and $\gamma \neq 1$ and for every $C_t > 0$, and including $C_t = 0$ if $\gamma < 1$, the equilibrium pricing equation can be expressed more explicitly as follows,

$$P_t = \beta \mathrm{E}_t[D_{t+1}^{1-\gamma} D_t^{\gamma} + D_{t+1}^{-\gamma} D_t^{\gamma} P_{t+1}]. \tag{4.1}$$

The model is closed by specifying a stochastic process for D_t,

$$\log D_t = (1-\rho)\log D^* + \rho \log D_{t-1} + \sigma \epsilon_t, \tag{4.2}$$

with given parameters $\rho \in (-1, 1)$, $D^* > 0$, and $\sigma > 0$; ϵ_t are independent and identically distributed (IID) standard Gaussian random variables; ϵ_t is unobservable at time $t-1$. We assume that the investor recognizes this stochastic process and the parameter values. By taking expectations of (4.2) and setting $\mathrm{E}[\log D_t] = \mathrm{E}[\log D_{t-1}]$, we find that $\log D^*$ is the long-run mean of $\log D_t$. We call D^* the *steady-state value* of D_t.

Taking expectations of (4.1) and evaluating D_t and D_{t+1} at D^*, we obtain

$$\begin{aligned} \mathrm{E}[P_t] &= \beta D^* + \beta \mathrm{E}\{\mathrm{E}_t[P_{t+1}]\} \\ &= \beta D^* + \beta \mathrm{E}[P_{t+1}]. \end{aligned}$$

Setting $\mathrm{E}[P_t] = \mathrm{E}[P_{t+1}]$, we obtain the long-run mean of P_t, denoted by P^*,

$$P^* = \frac{\beta}{1-\beta} D^*. \tag{4.3}$$

We also call P^* the steady-state value of P_t.

4.2.2 Log-Linearized Economic System

It is useful to work with log-linear approximations of (4.1)-(4.2) because we can obtain closed-form solutions for the quantities of interested in this article. The strategy is to use a first-order Taylor approximation around the steady states to replace the equations with approximations, which are linear in the log-deviations of the variables. We now define the variables $d_t \equiv \log D_t - \log D^*$ and $p_t \equiv \log P_t - \log P^*$. Following the approximation rules by Uhlig (2001), we have

$$D_t = D^* \exp(d_t) \Rightarrow D_t \approx D^*(1 + d_t), \tag{4.4}$$

$$P_t = P^* \exp(p_t) \Rightarrow P_t \approx P^*(1 + p_t). \tag{4.5}$$

Furthermore, we have the following approximations,

$$D_{t+1}^{1-\gamma} D_t^{\gamma} \approx D^*[1 + (1-\gamma)d_{t+1} + \gamma d_t], \tag{4.6}$$

$$D_{t+1}^{-\gamma} D_t^{\gamma} P_{t+1} \approx P^*(1 - \gamma d_{t+1} + \gamma d_t + p_{t+1}). \tag{4.7}$$

Using (4.4), we log-linearize the dividend model (4.2) as follows,

$$d_{t+1} = \rho d_t + \sigma \epsilon_{t+1}. \tag{4.8}$$

Using (4.5)-(4.7), we log-linearize the equilibrium price equation (4.1) as follows,

$$\begin{aligned} P^*(1+p_t) &= \beta \mathrm{E}_t\{D^*[1 + (1-\gamma)d_{t+1} + \gamma d_t] + P^*(1 - \gamma d_{t+1} + \gamma d_t + p_{t+1})\} \\ 1 + p_t &= \beta \mathrm{E}_t \left\{ \frac{D^*}{P^*}[1 + (1-\gamma)d_{t+1} + \gamma d_t] + (1 - \gamma d_{t+1} + \gamma d_t + p_{t+1}) \right\}. \end{aligned}$$

Substituting

$$\frac{D^*}{P^*} = \frac{1-\beta}{\beta},$$

implied by (4.3), and collecting term, we obtain

$$p_t = \gamma d_t + (1 - \beta - \gamma)\mathrm{E}_t[d_{t+1}] + \beta \mathrm{E}_t[p_{t+1}].$$

Further substituting $\mathrm{E}_t[d_{t+1}] = \rho d_t$, implied by (4.8), after simplification, we obtain the log-linearized equilibrium pricing equation,

$$p_t = \delta d_t + \beta \mathrm{E}_t[p_{t+1}], \tag{4.9}$$

where $\delta \equiv (1 - \beta - \gamma)\rho + \gamma$.

For the rest of this article, we base our analysis on the log-linearized economic system characterized by equations (4.8)-(4.9), which we summarize as follows for convenience,

$$p_t = \delta d_t + \beta \mathrm{E}_t[p_{t+1}],$$

$$d_{t+1} = \rho d_t + \sigma \epsilon_{t+1}.$$

This log-linearized system is also considered in Carceles-Poveda and Giannitsarou (2008) but we study different asset pricing issues. Henceforth, we simply refer to these equations as the pricing equation and the dividend model without emphasizing the term "log-linearized".

4.3 Equilibrium Analysis

4.3.1 A Benchmark Rational Expectations Equilibrium

Suppose that the representative investor's *perceived law of motion* (PLM) for p_t is given by

$$p_t = \mu_0 + \mu_1 d_t, \tag{4.10}$$

which, as we show later, is consistent with the RE solution of the model; μ_0 and μ_1 are parameters. Updating by one period and substituting in the dividend model (4.8), his PLM for p_{t+1} can be alternatively represented by

$$p_{t+1} = \mu_0 + \mu_1 \rho d_t + \mu_1 \sigma \epsilon_{t+1}. \tag{4.11}$$

Thus, his forecast of p_{t+1}, conditional on his time-t information set, denoted by $\hat{\mathrm{E}}_t[p_{t+1}]$, is given by

$$\hat{\mathrm{E}}_t[p_{t+1}] = \mu_0 + \mu_1 \rho d_t.$$

We can loosely interpret this investor as a *fundamental investor* because he formulates forecasts of future prices based on fundamental factors like dividends.

Substituting $\mathrm{E}_t[p_{t+1}]$ in the pricing equation (4.9) by $\hat{\mathrm{E}}_t[p_{t+1}]$, we obtain the *actual law of motion* (ALM) for p_t,

$$\begin{aligned} p_t &= \delta d_t + \beta(\mu_0 + \mu_1 \rho d_t) \\ &= \mu_0 \beta + (\delta + \beta \mu_1 \rho) d_t. \end{aligned} \tag{4.12}$$

A *rational expectations equilibrium* (REE) prevails when the PLM coincides with the ALM (*i.e.*, the functional forms and the coefficents of the PLM and the ALM are the identical). We formulate the REE in terms of a fixed point of the mapping from the PLM to the ALM defined by

$$T \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} = \begin{pmatrix} \mu_0 \beta \\ \delta + \beta \mu_1 \rho \end{pmatrix}. \tag{4.13}$$

Such a function T is sometimes called a T-map.

Therefore, the REE is given by

$$p_t = \overline{\mu}_0 + \overline{\mu}_1 d_t, \tag{4.14}$$

where

$$\overline{\mu}_0 = 0 \quad \text{and} \quad \overline{\mu}_1 = \frac{\delta}{1 - \beta\rho}. \tag{4.15}$$

4.3.2 Least-Squares Learning and Expectational Stability

Although the REE is well-defined, we now study whether it is *learnable* in the following sense. Suppose the representative investor believes that p_{t+1} follow the process (4.11), reproduced here for convenience,

$$p_{t+1} = \mu_0 + \mu_1 \rho d_t + \mu_1 \sigma \epsilon_{t+1}$$

but that μ_0 and μ_1 are unknown to him. The investor has data on the economy from periods $j = 0, 1, \ldots, t$. His time-t information set is $\{(p_j, d_j)\}_{j=0}^{t}$. We assume that the investor estimates μ_0 and μ_1 by a least squares regression of p_{t+1} on ρd_t and an intercept. Their estimates will be updated over time as more information is collected. Denote the estimates through time t by $(\mu_{0,t}, \mu_{1,t})$ and so his forecast of p_{t+1} is

$$\hat{\mathrm{E}}_t[p_{t+1}] = \mu_{0,t} + \mu_{1,t} \rho d_t. \tag{4.16}$$

The standard least squares formula gives the equations

$$\begin{bmatrix} \mu_{0,t} \\ \mu_{1,t} \end{bmatrix} = \left[\sum_{j=1}^{t} z_{j-1} z_{j-1}^{\top} \right]^{-1} \left[\sum_{j=1}^{t} z_{j-1} p_j \right], \tag{4.17}$$

where $z_j = (1, \rho d_j)^{\top}$.

The equilibrium price is now jointly defined by equations (4.9), (4.16), and (4.17). The question of interest is whether $(\mu_{0,t}, \mu_{1,t}) \to (\overline{\mu}_0, \overline{\mu}_1)$ as $t \to \infty$. If so, we say that the REE is learnable by least squares. In this article, we only consider learning by least squares so from now we simply use "learnable" to mean "learnable by least squares" for brevity.

Evans and Honkapohja (2001) show that learning problem of this type, under fairly general assumptions, the estimates converge to the REE if and only if certain stability conditions, known as *expectational stability* (E-stability) conditions, are satisfied. Given the mapping (4.13), E-stability of $(\overline{\mu}_0, \overline{\mu}_1)$ is defined as local asymptotic stability of the ordinary differential equation (ODE)

$$\frac{\mathrm{d}}{\mathrm{d}s}(\mu_0, \mu_1) = T(\mu_0, \mu_1) - (\mu_0, \mu_1)$$

at $(\overline{\mu}_0, \overline{\mu}_1)$, where s denotes notional time. Here, we omit the argument of $(\mu_0, \mu_1)(s)$ and simply use (μ_0, μ_1) for brievity. E-stability conditions are obtained by computing the Jacobian matrix of

$$h(\mu_0, \mu_1) \equiv T(\mu_0, \mu_1) - (\mu_0, \mu_1),$$

denoted by $\mathrm{D}h(\mu_0, \mu_1)$. If all of the eigenvalues of $\mathrm{D}h(\mu_0, \mu_1)$ evaluated at $(\overline{\mu}_0, \overline{\mu}_1)$ have negative real parts, then we say that the REE is E-stable or learnable, meaning that $(\mu_{0,t}, \mu_{1,t}) \to (\overline{\mu}_0, \overline{\mu}_1)$ as $t \to \infty$ with probability one.

For our current example, it can be shown that

$$\mathrm{D}h(\overline{\mu}_0, \overline{\mu}_1) = \begin{pmatrix} \beta - 1 & 0 \\ 0 & \beta\rho - 1 \end{pmatrix},$$

so both eigenvalues, the diagonal elements for a triangular matrix, are negative. This confirms the convergence to the REE.

4.3.3 A Self-Confirming Equilibrium

From this section onwards, we deviate from the standard benchmark model above in different ways to allow for investors' using technical analysis and study the corresponding market equilibria. To the best of our knowledge, the subsequent theoretical results are new.

Consider an exponentially weighted *moving average* defined by

$$a_t \equiv \alpha \sum_{j=-\infty}^{t} (1-\alpha)^{t-j} p_{j-1},$$

where $\alpha \in (0,1)$ is a given parameter controlling the window size[2]. We note that

$$\begin{aligned} a_{t+1} &= \alpha \sum_{j=-\infty}^{t+1} (1-\alpha)^{t+1-j} p_{j-1} \\ &= \alpha \sum_{j=-\infty}^{t} (1-\alpha)^{t+1-j} p_{j-1} + \alpha p_t \\ &= (1-\alpha)\alpha \sum_{j=-\infty}^{t} (1-\alpha)^{t-j} p_{j-1} + \alpha p_t \\ &= (1-\alpha) a_t + \alpha p_t. \end{aligned}$$

Since a_{t+1} is a weighted average of a_t and p_t, this quantity is known given the time-t information set.

Now, suppose that the representative investor's PLM for p_t is given by

$$p_t = \pi_0 + \pi_d d_t + \pi_a a_t, \tag{4.18}$$

where π_0, π_d, and π_a are parameters. That is, the investor overparametrizes his forecasting model by including a_t. One may wonder why the investor would want to include a_t. An economic rationale is that the investor is aware that some other

[2] The choice of exponentially moving averages, instead of simple moving averages used in previous chapters, is also commonly considered in the literature. However, the former can be viewed as a good approximation of the latter, or *vice versa*.
In practice, the industry convention is to set $\alpha = \frac{2}{N+1}$ to compute the N-day moving average.

investors are referring to this technical indicator. Hence, there is good reason to believe that prices are at least partially affected by a_t. From an econometric perspective, the investor will use the least-squares approach to determine the optimal value of π_a, and thus there is no disadvantage to include this additional variable—the estimate of π_a could be simply zero.

Note also that an equivalent specification is $p_t = \psi_0 + \psi_d d_t + \psi_x x_t$ with $x_t \equiv a_t - p_t$ being a momentum indicator measuring the percentage difference between the moving average and the price. This is a special case of double moving averages when the shorter one is simply the current price. Solving for p_t and reparametrizing give us back the specification (4.18).

Updating by one period, his PLM for p_{t+1} can be alternatively represented by

$$p_{t+1} = \pi_0 + \pi_d \rho d_t + \pi_a[(1-\alpha)a_t + \alpha p_t] + \pi_d \sigma \epsilon_{t+1}. \tag{4.19}$$

Thus, his forecast of p_{t+1}, conditional on his time-t information set, denoted by $\hat{\mathrm{E}}_t[p_{t+1}]$, is given by

$$\hat{\mathrm{E}}_t[p_{t+1}] = \pi_0 + \pi_d \rho d_t + \pi_a(1-\alpha)a_t + \pi_a \alpha p_t.$$

Substituting $\mathrm{E}_t[p_{t+1}]$ in the pricing equation (4.9) by $\hat{\mathrm{E}}_t[p_{t+1}]$, we obtain the ALM for p_t,

$$p_t = \delta d_t + \beta[\pi_0 + \pi_d \rho d_t + \pi_a(1-\alpha)a_t + \pi_a \alpha p_t].$$

Rearranging and collecting terms, we obtain

$$p_t = \frac{\pi_0 \beta}{1-\alpha\beta\pi_a} + \frac{\delta + \beta\pi_d\rho}{1-\alpha\beta\pi_a} d_t + \frac{(1-\alpha)\beta\pi_a}{1-\alpha\beta\pi_a} a_t. \tag{4.20}$$

Comparing (4.18) and (4.20), we can formulate the equilibrium in terms of a fixed point of the mapping from the PLM to the ALM defined by

$$T\begin{pmatrix}\pi_0\\ \pi_d\\ \pi_a\end{pmatrix}=\begin{pmatrix}\frac{\pi_0\beta}{1-\alpha\beta\pi_a}\\ \frac{\delta+\beta\pi_d\rho}{1-\alpha\beta\pi_a}\\ \frac{(1-\alpha)\beta\pi_a}{1-\alpha\beta\pi_a}\end{pmatrix}. \tag{4.21}$$

There are two fixed points, denoted by $(\overline{\pi}_0, \overline{\pi}_d, \overline{\pi}_a)$, for this mapping. The first fixed point is $(\overline{\mu}_0, \overline{\mu}_1, 0)$, which coincides with that of our benchmark REE. The second fixed point is

$$\overline{\pi}_0 = 0, \quad \overline{\pi}_d = \frac{\delta}{\beta(1-\alpha-\rho)}, \quad \text{and} \quad \overline{\pi}_a = \frac{1-\beta(1-\alpha)}{\alpha\beta}. \tag{4.22}$$

This identifies two possible equilibria in this scenario.

We call the second solution a self-confirming equilibrium in that sense that the investor believes the data generating process also depends on the moving averages, and because he believes so, it turns out to be true for the above parameters. In the next section, we will examine whether this equilibrium is worth studying.

Given the form of the equilibrium ALM for p_{t+1},

$$p_{t+1} = \overline{\pi}_0 + \overline{\pi}_d d_{t+1} + \overline{\pi}_a a_{t+1},$$

and the investor's optimal forecast of p_{t+1},

$$\hat{\mathrm{E}}_t[p_{t+1}] = \overline{\pi}_0 + \overline{\pi}_d \rho d_t + \overline{\pi}_a(1-\alpha)a_t + \overline{\pi}_a \alpha p_t,$$

we note that the forecast errors, defined by $p_{t+1} - \hat{\mathrm{E}}_t[p_{t+1}]$, are $\overline{\pi}_d \sigma \epsilon_{t+1}$, which are IID random variables. To see this, we have

$$\begin{aligned} p_{t+1} - \hat{\mathrm{E}}_t[p_{t+1}] &= \overline{\pi}_0 + \overline{\pi}_d d_{t+1} + \overline{\pi}_a a_{t+1} - [\overline{\pi}_0 + \overline{\pi}_d \rho d_t + \overline{\pi}_a (1-\alpha) a_t + \overline{\pi}_a \alpha p_t] \\ &= \overline{\pi}_d (d_{t+1} - \rho d_t) + \overline{\pi}_a \{a_{t+1} - [(1-\alpha) a_t + \alpha p_t]\} \\ &= \overline{\pi}_d \sigma \epsilon_{t+1}. \end{aligned}$$

The significance of this result is that the representative investor does not make systematic forecast errors in either of these two equilibria.

4.3.4 Learnability of the Self-Confirming Equilibrium

Now, suppose the representative investor believes that p_{t+1} follows the process (4.19), reproduced here for convenience,

$$p_{t+1} = \pi_0 + \pi_d \rho d_t + \pi_a a_{t+1} + \pi_d \sigma \epsilon_{t+1},$$

but that π_0, π_d, and π_a are unknown to him. His time-t information set is $\{(p_j, d_j, a_{j+1})\}_{j=0}^{t}$. We assume that the investor estimates π_0, π_d, and π_a by a least squares regression of p_{t+1} on ρd_t, a_{t+1}, and an intercept. Their estimates will be updated over time as more information is collected. Denote the estimates through time t by $(\pi_{0,t}, \pi_{d,t}, \pi_{a,t})$ and so his forecast of p_{t+1} is

$$\hat{\mathrm{E}}_t[p_{t+1}] = \pi_{0,t} + \pi_{d,t} \rho d_t + \pi_{a,t} a_{t+1}. \tag{4.23}$$

The standard least-squares formula gives the equations

$$\begin{bmatrix} \pi_{0,t} \\ \pi_{d,t} \\ \pi_{a,t} \end{bmatrix} = \left[\sum_{j=1}^{t} z_{j-1} z_{j-1}^{\top} \right]^{-1} \left[\sum_{j=1}^{t} z_{j-1} p_j \right], \tag{4.24}$$

where $z_j = (1, \rho d_j, a_{j+1})^\top$.

The equilibrium price is now jointly defined by equations (4.9), (4.23), and (4.24). We have already shown that the REE is E-stable. Indeed, in this case of overparameterization, we say that this REE is *strongly E-stable*. It remains to examine whether it is also true for the second equilibrium. That is, whether $(\pi_{0,t}, \pi_{d,t}, \pi_{a,t}) \to (\overline{\pi}_0, \overline{\pi}_d, \overline{\pi}_a)$, given by (4.22), as $t \to \infty$.

Given the mapping (4.21), the E-stability of $(\overline{\pi}_0, \overline{\pi}_d, \overline{\pi}_a)$ is defined as local asymptotic stability of the ODE

$$\frac{\mathrm{d}}{\mathrm{d}s}(\pi_0, \pi_d, \pi_a) = T(\pi_0, \pi_d, \pi_a) - (\pi_0, \pi_d, \pi_a)$$

at $(\overline{\pi}_0, \overline{\pi}_d, \overline{\pi}_a)$, where s denotes notional time. E-stability conditions are obtained by computing the Jacobian matrix of

$$h(\pi_0, \pi_d, \pi_a) \equiv T(\pi_0, \pi_d, \pi_a) - (\pi_0, \pi_d, \pi_a),$$

denoted by $\mathrm{D}h(\pi_0, \pi_d, \pi_a)$. The eigenvalues of $\mathrm{D}h(\pi_0, \pi_d, \pi_a)$ are

$$-\frac{\beta}{\alpha\beta\pi_a - 1} - 1, \quad -\frac{\beta\rho}{\alpha\beta\pi_a - 1} - 1, \quad \text{and} \quad \frac{\beta(\alpha-1)}{\alpha\beta\pi_a - 1} - \frac{\alpha\beta^2\pi_a(\alpha-1)}{(\alpha\beta\pi_a - 1)^2} - 1.$$

Evaluating these eigenvalues at $(\overline{\pi}_0, \overline{\pi}_d, \overline{\pi}_a)$, given by (4.22), we have

$$\frac{\alpha}{1-\alpha}, \quad \frac{\alpha+\rho-1}{1-\alpha}, \quad \text{and} \quad \frac{1-\beta(1-\alpha)}{\beta(1-\alpha)}.$$

Since the first eigenvalue, $\alpha/(1-\alpha)$, with $\alpha \in (0,1)$, is strictly positive. It suffices to conclude that the second equilibrium is E-unstable, that is, $(\pi_{0,t}, \pi_{d,t}, \pi_{a,t}) \to (\overline{\pi}_0, \overline{\pi}_d, \overline{\pi}_a)$ as $t \to \infty$ with probability zero. Hence, the economy will not converge to the self-confirming equilibrium, and thus we can conclude that this equilibrium is not worth studying.

We summarize the main findings of Sections 4.3.3-4.3.4 by the following proposition.

Proposition 4.1. *Suppose that the representative investor's PLM for p_t is given by $p_t = \pi_0 + \pi_d d_t + \pi_a a_t$ as in* (4.18). *There exist two possible equilibria, one coincides with the REE with $\pi_a = 0$ and the other is a self-confirming equilibrium with $\pi_a \neq 0$. The first equilibrium is least-squares learnable but not the second one.*

4.3.5 Technical Forecasting with Moving Averages

Let us consider a measure-zero *technical investor* interpreted as a minority with no impact on the economic system, *i.e.*, not the representative investor. We assume that this technical investor ignores fundamentals (dividends) but forecasts stock prices purely based on moving averages of past prices. Specifically, his PLM for p_t is given by

$$p_t = \tau_0 + \tau_1 a_t + \omega e_t, \tag{4.25}$$

where τ_0, τ_1, and ω are parameters; e_t are errors with zero mean and unit variance, which need not be ϵ_t in the dividend model (4.8). It is equivalent to think of this technical investor as an outside observer, say an econometrician, who regresses stock prices on moving averages and an intercept.

Suppose that the stock prices are generated according to the unique stable REE. That is, the ALM for p_t is given by (4.14)-(4.15). In this case, the technical investor's PLM for p_t does not coincide with the ALM, *i.e*, his forecasting model is *misspecified.*

Although the technical investor is using a misspecified forecasting model, we require that he does not make systematic forecast errors in the sense that (τ_0, τ_1) satisfy the standard least-squares orthogonality conditions:

$$\begin{aligned} \mathrm{E}[\omega e_t] &= \mathrm{E}[p_t - \tau_0 - \tau_1 a_t] = 0, \\ \mathrm{E}[a_t \omega e_t] &= \mathrm{E}[a_t(p_t - \tau_0 - \tau_1 a_t)] = 0. \end{aligned} \tag{4.26}$$

That is, we require $\tau_0 + \tau_1 a_t$ to be an unbiased predictor of p_t and the forecast error be uncorrelated with the predictive variable a_t.

The solutions for (τ_0, τ_1, ω) are then

$$\overline{\tau}_0 = \mathrm{E}[p_t] - \overline{\tau}_1 \mathrm{E}[a_t], \tag{4.27}$$

$$\overline{\tau}_1 = \frac{\mathrm{E}[p_t a_t] - \mathrm{E}[p_t]\mathrm{E}[a_1]}{\mathrm{E}[a_t^2] - (\mathrm{E}[a_t])^2}, \tag{4.28}$$

and

$$\overline{\omega}^2 = \mathrm{E}[(p_t - \overline{\tau}_0 - \overline{\tau}_1 a_t)^2], \tag{4.29}$$

which are explicit up to some moment equations of economic variables.

To compute the moment equations required in (4.27)-(4.29), it is convenient to express the economic system (p_t, d_t, a_t) as a VAR,

$$\begin{bmatrix} p_t \\ d_t \\ a_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & \overline{\mu}_1 \rho & 0 \\ 0 & \rho & 0 \\ \alpha & 0 & 1-\alpha \end{bmatrix} \begin{bmatrix} p_{t-1} \\ d_{t-1} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} \overline{\mu}_1 \sigma \\ \sigma \\ 0 \end{bmatrix} \epsilon_t.$$

More compactly, we can write

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \Sigma \epsilon_t, \tag{4.30}$$

where $y_t \equiv (p_t, d_t, a_t)^\top$ and the definitions of Φ_0, Φ_1, and Σ follow directly from the VAR model. It may look redundant to keep Φ_0 but we shall refer to this general VAR setting later.

Since the eigenvalues of Φ_1, namely 0, ρ, and $1-\alpha$, are all within the unit circle, this VAR is covariance-stationary, meaning there exist a 3×1 constant vector $\overline{y}$ and a 3×3 constant symmetric matrix C_y such that $\mathrm{E}[y_t] = \overline{y}$ and $\mathrm{E}[y_t y_t^\top] = C_y$.

To find the stationary mean $\overline{y}$, we take expectations of (5.14) to obtain

$$\begin{aligned} \mathrm{E}[y_{t+1}] &= \Phi_0 + \Phi_1 \mathrm{E}[y_t] \\ \overline{y} &= \Phi_0 + \Phi_1 \overline{y}. \end{aligned}$$

Solving for $\overline{y}$, we obtain

$$\mathrm{E}[y_t] = \overline{y} = (I - \Phi_1)^{-1} \Phi_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

where I is a 3×3 identity matrix.

To find the stationary variance-covariance matrix C_y, we post-multiply (5.14) by

$y_{t+1}^\top$ to obtain

$$\begin{aligned} y_{t+1}y_{t+1}^\top &= \Phi_0 y_{t+1}^\top + \Phi_1 y_t y_{t+1}^\top + \Sigma\epsilon_{t+1}y_{t+1}^\top \\ &= \Phi_0 y_{t+1}^\top + \Phi_1 y_t(\Phi_0^\top + y_t^\top \Phi_1^\top + \Sigma\epsilon_{t+1}) + \Sigma^\top \epsilon_{t+1}(\Phi_0^\top + y_t^\top \Phi_1^\top + \Sigma\epsilon_{t+1}). \end{aligned}$$

Taking expectations, we have

$$C_y = \Phi_0 \overline{y}^\top + \Phi_1 \overline{y}\Phi_0^\top + \Phi_1 C_y \Phi_1^\top + \Sigma\Sigma^\top,$$

which can be expressed as the following *Lyapunov equation*,

$$\Phi_1 C_y \Phi_1^\top - C_y + Q = 0,$$

where

$$Q \equiv \Phi_0 \overline{y}^\top + \Phi_1 \overline{y}\Phi_0^\top + \Sigma\Sigma^\top.$$

We can explicitly solve for C_y by using

$$(I - \Phi_1 \otimes \Phi_1)\text{vec}(C_y) = \text{vec}(Q),$$

where I is a conformable identity matrix, $\otimes$ is the Kronecker product operator, and $\text{vec}(\cdot)$ is the vectorization operator. We first solve for $\text{vec}(C_y)$ by solving the linear equations and then extract $C_{y,1,1} = \text{E}[p_t^2]$, $C_{y,1,3} = \text{E}[p_t a_t]$, and $C_{y,3,3} = \text{E}[a_t^2]$ to compute $(\overline{\tau}_0, \overline{\tau}_1, \overline{\omega})$ given by (4.27)-(4.29). We present the solutions of the required moments as follows.

$$\text{E}[p_t^2] = \frac{\overline{\mu}_1^2 \sigma^2}{1-\rho^2},$$

$$\text{E}[p_t a_t] = \frac{\alpha\rho\overline{\mu}_1^2\sigma^2}{(1-\rho^2)(1+(\alpha-1)\rho)},$$

$$\mathrm{E}[a_t^2] = \frac{\alpha \overline{\mu}_1^2 (1-(\alpha-1)\rho)\sigma^2}{(2-\alpha)(1-\rho^2)(1+(\alpha-1)\rho)}.$$

After some algebra, we obtain

$$\overline{\tau}_0 = 0, \quad \overline{\tau}_1 = \frac{(2-\alpha)\rho}{1+(1-\alpha)\rho}, \quad \text{and} \quad \overline{\omega}^2 = \kappa \overline{\mu}_1^2 \sigma^2,$$

where

$$\kappa = \frac{1}{1-(1-\alpha)^2\rho^2} > 1.$$

We note that if $\rho \in (0,1)$ then $\overline{\tau}_1 \in (0,1)$. Note also that $\overline{\mu}_1^2\sigma^2$ is the variance of forecast error implied by the fundamental investor's forecasting model (4.11). Since $\kappa > 1$, the technical investor is exposed to a higher variance of forecast error than that of a fundamental investor.

Now, suppose the technical investor does not know the values of (τ_0, τ_1) and has to estimate them by a least squares regression of p_t on a_t and an intercept. Suppose also that he updates the estimates over time as more information is collected. By the stationarity of the VAR, we know that the least squares estimates $(\tau_{0,t}, \tau_{1,t})$ converge to $(\overline{\tau}_0, \overline{\tau}_1)$ as $t \to \infty$ with probability one.

We summarize the main finding of this section by the following proposition.

Proposition 4.2. *Suppose stock prices are generated according to the stable REE as given by* (4.14)-(4.15). *Suppose also that a measure-zero technical investor's PLM for p_t is given by $p_t = \tau_0 + \tau_1 a_t + \omega e_t$ as in* (4.25). *There exist statistically optimal values (τ_0, τ_1) in the sense that they satisfy the least-squares orthogonality conditions* (4.26).

This proposition states that although the technical investor's forecasting model is inconsistent with the actual stock price model, he would still find that moving averages of past prices are useful to predict future prices. Similarly, when an outside observer, say an econometrician, would like to assess the predictive power of moving averages of past prices on future prices by a regression, he would find strong evidence supporting such predictive power if he did not include dividend in the regression as a control.

4.3.6 Technical Forecasting and Two Restricted Perceptions Equilibria

We now consider the case that the representative investor is a technical investor who has a PLM for p_t represented by (4.25), reproduced here for convenience,

$$p_t = \tau_0 + \tau_1 a_t + \omega e_t,$$

where τ_0, τ_1, and ω are parameters; e_t are errors with zero mean and unit variance, which need not be ϵ_t in the dividend model (4.8). We once again require (τ_0, τ_1) to satisfy the moment conditions (4.26). The difference here is that the investor's PLM now affects the ALM for p_t.

Updating the PLM by one period and taking conditional expectations, we obtain

$$\begin{aligned}\hat{\mathrm{E}}_t[p_{t+1}] &= \tau_0 + \tau_1 a_{t+1} \\ &= \tau_0 + \tau_1(1-\alpha)a_t + \tau_1 \alpha p_t.\end{aligned}$$

Substituting $\mathrm{E}_t[p_{t+1}]$ in the pricing equation (4.9) by $\hat{\mathrm{E}}_t[p_{t+1}]$, we obtain,

$$p_t = \delta d_t + \beta[\tau_0 + \tau_1(1-\alpha)a_t + \tau_1 \alpha p_t].$$

Rearranging and collecting terms, we obtain the ALM for p_t,

$$p_t = \frac{\tau_0 \beta}{1-\beta\tau_1\alpha} + \frac{\delta}{1-\beta\tau_1\alpha} d_t + \frac{\beta\tau_1(1-\alpha)}{1-\beta\tau_1\alpha} a_t. \tag{4.31}$$

Further substituting the ALM (4.31) into the moment conditions (4.26), the solutions of (τ_0, τ_1) solve the following simultaneous equations,

$$\left(\frac{\tau_0\beta}{1-\beta\tau_1\alpha} - \tau_0\right) + \left(\frac{\beta\tau_1(1-\alpha)}{1-\beta\tau_1\alpha} - \tau_1\right) \mathrm{E}[a_t] = 0, \tag{4.32}$$

and

$$\left(\frac{\tau_0\beta}{1-\beta\tau_1\alpha} - \tau_0\right) \mathrm{E}[a_t] + \frac{\delta}{1-\beta\tau_1\alpha} \mathrm{E}[d_t a_t] + \left(\frac{\beta\tau_1(1-\alpha)}{1-\beta\tau_1\alpha} - \tau_1\right) \mathrm{E}[a_t^2] = 0, \tag{4.33}$$

where the moments $\mathrm{E}[a_t]$, $\mathrm{E}[d_t a_t]$, and $\mathrm{E}[a_t^2]$ are also functions of (τ_0, τ_1) to be determined as follows.

We expand the right-hand side of the ALM (4.31) as follows.

$$p_t = \frac{\tau_0\beta}{1-\beta\tau_1\alpha} + \frac{\delta}{1-\beta\tau_1\alpha}(\rho d_{t-1} + \sigma\epsilon_t) + \frac{\beta\tau_1(1-\alpha)}{1-\beta\tau_1\alpha}[(1-\alpha)a_{t-1} + \alpha p_{t-1}].$$

Collecting terms, we obtain

$$\begin{aligned} p_t = {} & \frac{\tau_0\beta}{1-\beta\tau_1\alpha} + \frac{\beta\tau_1(1-\alpha)\alpha}{1-\beta\tau_1\alpha} p_{t-1} + \frac{\delta\rho}{1-\beta\tau_1\alpha} d_{t-1} + \frac{\beta\tau_1(1-\alpha)^2}{1-\beta\tau_1\alpha} a_{t-1} \\ & + \frac{\delta\sigma}{1-\beta\tau_1\alpha}\epsilon_t. \end{aligned}$$

To compute the moment equations required in (4.32)-(4.33), we once again summarize the economic system (p_t, d_t, a_t) by a VAR as in (5.14) with the following coefficients,

$$\Phi_0 = \begin{bmatrix} \frac{\tau_0 \beta}{1-\beta\tau_1\alpha} \\ 0 \\ 0 \end{bmatrix}, \Phi_1 = \begin{bmatrix} \frac{\beta\tau_1(1-\alpha)\alpha}{1-\beta\tau_1\alpha} & \frac{\delta\rho}{1-\beta\tau_1\alpha} & \frac{\beta\tau_1(1-\alpha)^2}{1-\beta\tau_1\alpha} \\ 0 & \rho & 0 \\ \alpha & 0 & 1-\alpha \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{\delta\sigma}{1-\beta\tau_1\alpha} \\ \sigma \\ 0 \end{bmatrix}. \tag{4.34}$$

Then, we follow a similar procedure as in Section 4.3.7 to obtain the expressions for the moments $\mathrm{E}[a_t]$, $\mathrm{E}[d_t a_t]$, and $\mathrm{E}[a_t^2]$. Substituting these moments into (4.32)-(4.33), we find two possible solutions for (τ_0, τ_1), each corresponds to a *restricted perceptions equilibrium* (RPE), characterized by

$$\overline{\tau}_0 = 0 \tag{4.35}$$

and

$$\overline{\tau}_1 = \frac{\alpha+\rho-1}{2\alpha} + \frac{1+(1-\alpha)\rho \pm \sqrt{\varphi}}{2\alpha\beta}, \tag{4.36}$$

where

$$\varphi \equiv (1-\alpha+\beta)^2\rho^2 + 2(1-\alpha-\beta)(1+(1-\alpha)\beta)\rho + (1-(1-\alpha)\beta)^2.$$

We now show that $\overline{\tau}_1$ is a real number by showing that φ is positive. By noting that φ is a convex quadratic function of ρ (for the moment, we take $\rho \in \mathbb{R}$), we find that the global minimum of φ is

$$\min_{\rho\in\mathbb{R}} \varphi = \frac{4\alpha\beta(1-\alpha)(2-\alpha)(1-\beta^2)}{(1-\alpha-\beta)^2} > 0.$$

Hence, we know that $\varphi > 0$ for any $\rho \in (-1, 1)$ and so the solution $\overline{\tau}_1$ is well-defined.

4.3.7 Learnability of the Two Restricted Perceptions Equilibria

Now, suppose the representative investor described in the previous section does not know the parameter values of (τ_0, τ_1) and uses the forecasting model

$$\hat{\mathrm{E}}_t[p_{t+1}] = \tau_{0,t} + \tau_{1,t}a_{t+1},$$

where $(\tau_{0,t}, \tau_{1,t})$ are least squares estimates, based on his time-t information set $\{(p_j, a_{j+1})\}_{j=0}^{t}$, given by

$$\begin{bmatrix} \tau_{0,t} \\ \tau_{1,t} \end{bmatrix} = \left[\sum_{j=1}^{t} z_j z_j^{\top} \right]^{-1} \left[\sum_{j=1}^{t} z_j p_j \right], \tag{4.37}$$

where $z_j = (1, a_j)^{\top}$.

Replacing (τ_0, τ_1) by $(\tau_{0,t}, \tau_{1,t})$ in equation (4.31), the ALM for p_t becomes

$$p_t = \frac{\tau_{0,t}\beta}{1 - \beta\tau_{1,t}\alpha} + \frac{\delta}{1 - \beta\tau_{1,t}\alpha} d_t + \frac{\beta\tau_{1,t}(1-\alpha)}{1 - \beta\tau_{1,t}\alpha} a_t. \tag{4.38}$$

The equilibrium price is now implicitly determined by equations (4.37)-(4.38) (note that $(\tau_{0,t}, \tau_{1,t})$ are functions of p_t). The question of interest is whether $(\tau_{0,t}, \tau_{1,t}) \to (\overline{\tau}_0, \overline{\tau}_1)$, any of the two RPEs given by (4.35)-(4.36), as $t \to \infty$.

To study the learning problem of this type, it is useful to first express (4.37) as the following *recursive least squares* (RLS) algorithm,

$$\tau_t = \tau_{t-1} + t^{-1} R_t^{-1} z_t (p_t - z_t^\top \tau_{t-1}), \tag{4.39}$$

$$R_t = R_{t-1} + t^{-1}(z_t z_t^\top - R_{t-1}), \tag{4.40}$$

where $\tau_t \equiv (\tau_{0,t}, \tau_{1,t})^\top$. We note that z_t has argument τ_{t-1}, R_t and p_t have argument τ_t.

Define $\tau_t \equiv (\tau_{0,t}, \tau_{1,t})$ and $\overline{\tau} \equiv (\overline{\tau}_0, \overline{\tau}_1)$. Following Evans and Honkapohja (2001), under appropriate regularity conditions, if $\tau_t \to \overline{\tau}$ as $t \to \infty$, then $\overline{\tau}$ is a stable fixed point of the following system of ODEs[3],

$$\frac{\mathrm{d}\tau}{\mathrm{d}s} = \mathrm{E}\{R^{-1}(\tau) z(\tau)[p(\tau) - z^\top(\tau)\tau]\},$$

$$\frac{\mathrm{d}R}{\mathrm{d}s} = \mathrm{E}[z(\tau) z^\top(\tau) - R(\tau)],$$

where s denotes notional time. Also, we use τ to denote $\tau(s)$ for brevity.

Setting $\mathrm{d}R/\mathrm{d}s = 0$, we have $R(\tau) = \mathrm{E}[z(\tau)z^\top(\tau)]$ for a fixed τ. Substituting $R^{-1}(\tau)$ into $\mathrm{d}\tau/\mathrm{d}s$ and equating to zero, with some simplification, we find that the local stability conditions are obtained by computing the Jacobian matrix of

$$h(\tau) \equiv \left(\mathrm{E}[z(\tau)z^\top(\tau)]\right)^{-1} \mathrm{E}[z(\tau)p(\tau)] - \tau, \tag{4.41}$$

denoted by $\mathrm{D}h(\tau)$, provided the expectations exit. If all of the eigenvalues of

[3] Technically speaking, the E-stability conditions discussed in the previous sections are also results of the convergence conditions for some RLS algorithms. However, the problem at hand does not admit a straightforward T-map as before. Therefore, we defer introducing the concept of RLS convergence but directly apply the stability results based on the T-maps in the previous sections.

D$h(\tau)$ evaluated at $\bar{\tau}$ have negative real parts, then we say $\bar{\tau}$ is a learnable RPE, meaning that $\tau_t \to \bar{\tau}$ as $t \to \infty$ with probability one. We note that for the expectations in $h(\bar{\tau})$ to exist, we require the all of the eigenvalues of $\Phi_1(\tau)$, given by (4.34), evaluated at $\bar{\tau}$ to be within the unit circle. It can be shown that the associated eigenvalues of $\Phi_1(\tau)$ are 0, ρ, and $(1-\alpha)/(1-\alpha\beta\tau_1)$. For the moment, let us assume that the following stationarity condition holds

$$\left|\frac{1-\alpha}{1-\alpha\beta\tau_1}\right| < 1. \tag{4.42}$$

Expressing (4.41) more explicitly as

$$h(\tau) = \begin{bmatrix} 1 & \mathrm{E}[a(\tau)] \\ \mathrm{E}[a(\tau)] & \mathrm{E}[a^2(\tau)] \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{E}[p(\tau)] \\ \mathrm{E}[p(\tau)a(\tau)] \end{bmatrix} - \tau,$$

where the moment equations can be computed using the VAR model (5.14) with coefficeints given by (4.34). It follows that the eigenvalues of $h(\tau)$ are

$$\frac{(1-\rho)\beta}{1-\alpha\beta\tau_1-(1-\alpha)\rho} - 1 \quad \text{and} \quad \frac{(1-\alpha)(1-\rho^2)\beta}{(1-\alpha\beta\tau_1-(1-\alpha)\rho)^2} - 1. \tag{4.43}$$

Let us denote the two possible solutions for $\bar{\tau}_1$ in (4.36) by

$$\bar{\tau}_1^+ = \frac{\alpha+\rho-1}{2\alpha} + \frac{1+(1-\alpha)\rho+\sqrt{\varphi}}{2\alpha\beta}$$

and

$$\bar{\tau}_1^- = \frac{\alpha+\rho-1}{2\alpha} + \frac{1+(1-\alpha)\rho-\sqrt{\varphi}}{2\alpha\beta}.$$

The eigenvalues (4.43) evaluated at $\overline{\tau}_1^+$ are

$$\lambda_1^+ \equiv \frac{2\beta(1-\rho)}{1+(1-\alpha)\beta+(1-\alpha-\beta)\rho-\sqrt{\varphi}} - 1$$

and

$$\lambda_2^+ \equiv \frac{4\beta(1-\alpha)(1-\rho^2)}{(1+(1-\alpha)\beta+(1-\alpha-\beta)\rho-\sqrt{\varphi})^2} - 1.$$

Similarly, evaluating the eigenvalues (4.43) at $\overline{\tau}_1^-$ gives

$$\lambda_1^- \equiv \frac{2\beta(1-\rho)}{1+(1-\alpha)\beta+(1-\alpha-\beta)\rho+\sqrt{\varphi}} - 1$$

and

$$\lambda_2^- \equiv \frac{4\beta(1-\alpha)(1-\rho^2)}{(1+(1-\alpha)\beta+(1-\alpha-\beta)\rho+\sqrt{\varphi})^2} - 1.$$

It can be verified that the $\lambda_2^+ \geq 0$ so $\overline{\tau}_1^+$ is not a learnable RPE. For $\overline{\tau}_1^-$, we find that it always satisfies the stationarity condition (4.42) with $\lambda_1^- < 0$. Thus, it remains to check whether $\lambda_2^- < 0$.

We find that for λ_2^- to be strictly negative while satisfying the VAR stationarity condition, we require any one of the following three conditions to be satisfied[4]:

(i) $-1 < \rho \leq -\frac{1}{2}$ or $\frac{1}{2} \leq \rho < 1$; or

(ii) $0 < \alpha < \frac{1+2\rho}{2(1+\rho)}$ and $0 < \beta < \frac{1+2(1-\alpha)\rho}{-4(1-\alpha)\rho^2+2\rho+2(1-\alpha)}$; or

(iii) $\frac{1+2\rho}{2(1+\rho)} \leq \alpha < 1$.

But since $\rho \in (-1, 1)$, condition (iii) implies that a sufficient condition for the

[4]These conditions can be obtained by using a symbolic mathematical computation program such as *Mathematica*.

second RPE to be learnable is $\alpha \geq \frac{3}{4}$. This implies that the investors have to use a relatively short lookback period to compute the moving averages (see Note 2). In order words, the market has to be relatively informationally efficient in the sense that remote past prices do not affect future prices.

We summarize the main findings of Sections 4.3.6-4.3.7 by the following proposition.

Proposition 4.3. *Suppose that the representative investor is a technical investor with the PLM for p_t given by $p_t = \tau_0 + \tau_1 a_t + \omega e_t$ as in (4.25). It is possible to have a RPE in which the investor cannot detect his misspecification but his expectations are otherwise optimal within a limited class of forecasting models. A sufficient condition for this RPE to be learnable is $\alpha \geq \frac{3}{4}$, i.e., a relatively short lookback period to compute the moving averages.*

This proposition states that it is possible for the stock market to reach a *stable* equilibrium even though all investors are technical investors who systematically use a misspecified forecasting model without recognizing their misspecification. To conclude, this example demonstrates that equilibrium stability need not require agents' rationality and thus looking beyond RE allows us to study other interesting equilibrium phenomena, as we attempt in this article.

4.4 Conclusion

While several empirical studies find evidence in favor of the usefulness of technical analysis, there is relatively limited theoretical justification. Several noisy rational expectations (RE) models demonstrate, with only a few explicitly built in the context of technical analysis, that past prices can contain useful information to predict future stock prices. However, as discussed in the literature, not only does RE implicitly make some rather strong assumptions (*e.g.* investors do no suffer from model misspecification, know the model parameters, and always form "perfect" expectations), there are potential difficulties to interpret the model results. For example, it is possible to have multiple equilibria but the RE assumption itself provides little guidance on selecting which is more plausible. Also, the existence and the uniqueness of a rational expectations equilibrium do not imply that investors can actually learn the equilibrium by standard adaptive learning rules, even with a large sample. This means that the equilibrium need not be robust to investors' making small forecast errors initially and neglects the notion that investors' expectations evolve over time.

Adapting Lucas' (1978) one-tree representative investor framework, this article replaces RE by adaptive learning (least-squares) and shows that moving averages of past stock prices can forecast future prices as an equilibrium phenomenon under different assumptions on the investor's forecasting model. Beyond the derivation and the characterization of equilibrium solutions, we also focus on proving equilibrium stability, *i.e.*, whether the investor can eventually learn the equilibrium.

This study is only an initial attempt to study the usefulness of technical analysis through an equilibrium model. One important difference to most noisy rational expectations models in the literature is that we do not consider an interactive multi-agent model with hierarchical information structure. For example, we could extend the model by allowing fundamental and technical investors to interact strategically. A challenging future task is to formulate an equilibrium under heterogeneous expectations and assess E-stability in this complex setting.

4.5 References

Barsky, R. B., & De Long, J. B. (1993). Why does the stock market fluctuate? Quarterly Journal of Economics, 108 (2), 291-311.

Blume, L., Easley, D., & O'hara, M. (1994). Market statistics and technical analysis: The role of volume. Journal of Finance, 49(1), 153-181.

Branch, W. A., & Evans, G. W. (2010). Asset return dynamics and learning. Review of Financial Studies, 23(4), 1651-1680.

Brennan, M. J., & Xia, Y. (2001). Stock price volatility and equity premium. Journal of Monetary Economics, 47(2), 249-283.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

Brown, D. P., & Jennings, R. H. (1989). On technical analysis. Review of Financial Studies, 2(4), 527-551.

Bullard, J., & Mitra, K. (2002). Learning about monetary policy rules. Journal of Monetary Economics, 49(6), 1105-1129.

Carceles-Poveda, E., & Giannitsarou, C. (2008). Asset pricing with adaptive learning. Review of Economic Dynamics, 11(3), 629-651.

Dothan, M. U., & Feldman, D. (1986). Equilibrium interest rates and multiperiod bonds in a partially observable economy. The Journal of Finance, 41(2), 369-382.

Evans, G. W. (1985). Expectational stability and the multiple equilibria problem in linear rational expectations models. Quarterly Journal of Economics, 1217-1233.

Evans, G. W., & Honkapohja, S. (2001). Learning and expectations in macroeconomics. Princeton University Press.

Evans, G. W., & Honkapohja, S. (2003). Expectations and the stability problem for optimal monetary policies. Review of Economic Studies, 70(4), 807-824.

Feldman, D. (2002). Production and the real rate of interest: A sample path equilibrium. European Finance Review, 6(2), 247-275.

Feldman, D. (2003). The Term Structure of Interest Rates: Bounded or Falling?. European Finance Review, 7(1), 103-113.

Feldman, David. (2007). Incomplete information equilibria: Separation theorems and other myths. Annals of Operations Research, 151(1), 119-149.

Grundy, B. D., & McNichols, M. (1989). Trade and the revelation of information through prices and direct disclosure. Review of Financial Studies, 2(4), 495-526.

Hansen, L. P., & Sargent, T. J. (1982). Instrumental variables procedures for estimating linear rational expectations models. Journal of Monetary Economics, 9(3), 263-296.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis, 48(5), 1433-1461.

Howitt, P. (1992). Interest rate control and nonconvergence to rational expectations. Journal of Political Economy, 776-800.

Lewellen, J., & Shanken, J. (2002). Learning, asset-pricing tests, and market efficiency. Journal of Finance, 57(3), 1113-1145.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. Econometrica, 46(6), 1429-1445.

Marcet, A., & Sargent, T. J. (1989a). Convergence of least-squares learning in environments with hidden state variables and private information. Journal of Political Economy, 97(6), 1306-1322.

Marcet, A., & Sargent, T. J. (1989b). Convergence of least squares learning mechanisms in self-referential linear stochastic models. Journal of Economic theory, 48(2), 337-368.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60(7), 1772-1791.

Pástor, L'., & Pietro, V. (2003). Stock valuation and learning about profitability. Journal of Finance, 58(5), 1749-1790.

Sargent, T. J. (2002). The conquest of American inflation. Princeton University Press.

Timmermann, A. (1994). Can agents learn to rational expectations? Some results on convergence and stability of learning in the UK stock market. The Economic Journal, 104(405), 777-797.

Timmermann, A. (1996). Excess volatility and predictability of stock prices in autoregressive dividend models with learning. The Review of Economic Studies, 63(4), 523-557.

Treynor, J. L., & Ferguson, R. (1985). In defense of technical analysis. Journal of Finance, 40(3), 757-773.

Uhlig, Harald. (2001). A toolkit for analysing non-linear dynamic stochastic models easily. Computational methods for the study of dynamic economies. Oxford University Press, 30-61.

Veronesi, P. (2000). How does information quality affect stock returns? Journal of Finance, 55(2), 807-837.

Wang, J. (1993). A model of intertemporal asset prices under asymmetric information. Review of Economic Studies, 60(2), 249-282.

Williams, N. (2004). Escape dynamics in learning models. Mimeo. Princeton University Press.

Zhou, G., & Zhu, Y. (2013). An equilibrium model of moving-average predictability and time-series momentum. Unpublished working paper, Washington University in St. Louis.

Chapter 5

Estimation of a Dynamic Model of Asset Prices and Technical Forecasting with Moving Averages

This article formulates a model to study the predictive power of moving averages as an equilibrium phenomenon with special attention to developing model estimation and testing strategies. Adapting Lucas' (1978) one-tree representative investor framework, prices are determined endogenously and are affected by the investor's forecasts of next period's price. By imposing different assumptions on the investor's forecasting model, we obtain several possible equilibria. A special feature of this model is that the parameters of the investor's forecasting model are also determined endogenously. Based on a set of orthogonality conditions implied by the equilibrium pricing equation and the investor's forecasting model, we develop estimation and testing strategies to examine whether the proposed equilibria are empirically supported.

5.1 Introduction

While several empirical studies find evidence for the predictive power of moving averages on stock prices, see, *e.g.* Brock, Lakonishok, and LeBaron (1992), Han, Yang, and Zhou (2013), and Neely *et al.* (2014), these studies are restricted to reduced-form estimation. The reduced-form approach does give us useful insights about the statistical relationship between economic variables but it provides limited guidance on whether such a relationship is an equilibrium phenomenon and how such a relationship is formed. One major difficulty is that the same (or similar) reduced-form model can be arrived at by economic models that impose very different assumptions on the agents' decision-making and actions. Another extreme is that the reduced-form equation being estimated cannot be generated by any plausible economic model.

As an attempt to take up this issue, this article formulates a structural model to study the predictive power of moving averages as an equilibrium phenomenon with special attention to developing model estimation and testing strategies. We begin by formulating an asset pricing model under Lucas' (1978) one-tree representative investor framework. Then, we propose several possible equilibria based on different assumptions on the investor's stock price forecasting model. Finally, we develop estimation and testing strategies to examine whether the proposed equilibria are empirically supported.

The equilibria considered in this article are labelled as either a rational expectations equilibrium (REE) or a restricted perceptions equilibrium (RPE). A REE prevails when the investor's forecasting model coincides with the equilibrium pricing equation (*i.e.*, with the same functional forms and parameters). We show that this is the case if the investor formulates forecasts of future prices based on dividends and observable shocks to dividends. A RPE prevails when the investor misspecifies his forecasting model but his forecasts are otherwise optimal within a limited class of forecasting models (see Evans and Honkapohja (2001) for a more detailed discussion about the concept of a RPE). We show that it is possible to arrive at a RPE when the investor ignores dividends but forecasts stock prices purely based on lagged prices or moving averages of past prices. Our model shows that the equilibrium pricing equation has the same or similar functional form under each of these equilibria. Therefore, reduced-form estimation of the regression relationship implied by the equilibrium pricing equation itself provides little guidance on selecting which equilibria is more plausible.

Using actual monthly data on dividends and stock index prices, we show that while the RPEs are theoretically attainable, they are empirically poorly supported in the sense that some of the model implied sample moment restrictions are systematically violated. By contrast, we find that the REE is more plausible with superior econometric performance.

While there is a large literature on heterogeneous agent models (HAMs) which also study how equilibrium prices are affected by investors' forecasts of future prices, the parameters of the investors' forecasting models are directly taken as either exogenous parameters or reduced-form estimates from data (see, *e.g.* Brock and Hommes, 1997, 1998; Boswijk *et al.* 2007; Hommes, 2013). In the standpoint of modelling strategy, our model differs from HAMs in that the parameters of the investor's forecasting model are determined *endogenously*, as functions of structural parameters representing time preference and risk aversion. The self-referential feature of the asset pricing model, namely prices are determined endogenously and affected by investor's forecasts of next period's price, allows us to pin down whether there exist plausible values of structural parameters such that the model as a whole can *simultaneously* explain both the observed prices and the parameters of the investor's forecasting model. The equilibrium pricing equation and the investor's forecasting model are then used to generate a set of population orthogonality conditions for us to develop an identification strategy based on the GMM approach proposed by Hansen and Singleton (1982). This feature of identification is also absent in HAMs. Our implementation is, however, different from Hansen and Singleton in that there is no arbitrary choice of "instrumental variables" because we make explicit assumptions about what information does the representative investor use (such that we can empirically identify his forecasting model).

Our model also relates to the literature on theoretical examination about the usefulness of technical analysis, see, *e.g.* Treynor and Ferguson (1985), Brown and Jennings (1989), Grundy and McNichols (1989), Blume, Easley, and O'Hara (1994), and Zhou and Zhu (2013). To the best of our knowledge, Zhou and Zhu is the only study presenting a model that explicitly illustrates the predictive power of moving averages as an equilibrium phenomenon. Their model follows the work by Wang (1993) and considers heterogenous investors. Using a different modelling strategy, we show that it is possible to admit equilibria, namely RPEs, in which all investors are identical trend followers or technical investors using moving averages. This suggests that allowing for heterogenous investors need not be a critical assumption to reach a market equilibrium. However, when we go one step further and formally test whether these equilibria are plausible with actual data, we do not find any empirical support.

This chapter can be viewed as an empirical version of Chapter 4. An important issue remains unanswered in the last chapter is that while we have identified two theoretically plausible equilibria, in the sense of learnability and long-run stability, the theory per se does not suggest which is a better description of the real world. As a result, an empirical investigation is called for.

However, the exact same model cannot be utilized directly because some aspects are simplified for a cleaner theoretical demonstration but not necessarily hold in real world. Therefore, some modifications and justifications are needed. First, the stationarity assumption on dividends (and so stock prices) clearly does not hold in the real world. Thus, we need to test dividends and prices are trend-stationary so that we can leave out their trends in the model with the understanding that investors can adjust the predictions from the model by accounting for their deterministic trends afterwards. Second, we find that detrended dividends do not follow a simple AR(1) process as first proposed but a more complicated ARMA(2,1) process. This fact has an immediately effect on all the equilibrium solutions obtained before and thus we have to rederive new solutions.

5.2 The Model

5.2.1 The Basic Setting

Consider a standard pure-exchange economy (Lucas, 1978) with a single consumer, interpreted as a *representative investor* "stand in" for a large number of identical investors. We assume that investors' opportunity set comprises a risky asset and a bond. Each unit of the stock held during period $t-1$, θ_t^s, yields a dividend payment D_t at time t; time-t stock prices are P_t. Each unit of bond held during period $t-1$, θ_t^b, yields a gross payment $1+r$ at time t; bond prices are normalized to one so the implied risk-free rate of return is r. The investor chooses the units to invest to the risky asset and the bond, (θ_t^s, θ_t^b), and consumption, C_t, in order to solve the life-time utility maximization problem:

$$\max_{\{C_t, \theta_t^s, \theta_s^b\}_{t=0}^{\infty}} \mathrm{E} \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma}, \quad \gamma > 0, \quad \gamma \neq 1,$$

subject to the budget constraint

$$(P_t + D_t)\theta_{t-1}^s + (1+r)\theta_{t-1}^b = C_t + P_t\theta_t^s + \theta_t^b,$$

where $\beta \in (0,1)$ is a discount factor, γ is the constant relative risk-aversion parameter, and E is an expectations operator.

In the equilibrium, we have $\theta_t^s = \theta_{t-1}^s$ and $\theta_t^b = \theta_{t-1}^b = 0$ such that $C_t = D_t$. It is a well-known result that under these equilibrium conditions the *equilibrium pricing equation* of this model is

$$P_t = \beta \mathrm{E}_t[D_{t+1}^{1-\gamma} D_t^{\gamma} + D_{t+1}^{-\gamma} D_t^{\gamma} P_{t+1}], \tag{5.1}$$

where the notation E_t indicates that the expectation is taken conditional on the investor's time-t information set (to be specified). Since the bond can be thought of an asset yielding a dividend $D_t = r$, where the bond price is always equal to one, the pricing equation (5.1) implies that the *equilibrium interest rate* of the model is

$$r = \frac{1-\beta}{\beta}. \tag{5.2}$$

The model is closed by specifying an exogenous stochastic process for D_t. It is given by

$$\log D_t = (1 - \phi_1 - \phi_2) \log D^* + \phi_1 \log D_{t-1} + \phi_2 \log D_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad (5.3)$$

an ARMA(2,1) model, where ε_t are independent and identically distributed (IID) exogenous shocks with mean zero and variance σ_ε^2; $(D^*, \phi_1, \phi_2, \theta_1, \sigma_\varepsilon^2)$ are given model parameters. We assume that the investor recognizes this stochastic process and the parameter values. Once D_t is observed at time t, the investor can immediately compute ε_t. His time-t information set is thus $\{(P_\tau, D_\tau, \varepsilon_\tau) : \tau \leq t\}$. Assuming stationary[1], it can be shown that $\log D^*$ is the long-run mean of $\log D_t$. We call D^* the *steady-state value* of D_t. Also, by the pricing equation (5.1), the implied steady-state value of P_t, denoted by P^*, is given by $P^* = D^*/r$.

5.2.2 Log-Linearized Economic System

Define $d_t \equiv \log D_t - \log D^*$ and $p_t \equiv \log P_t - \log P^*$. We can interpret (d_t, p_t) as percentage deviations from their steady-state values. Henceforth, we simply refer them to as log-dividend and log-price deviations. Applying the approximation rules by Uhlig (2001), we can log-linearize (5.1) and (5.3) as follows

$$p_t = \gamma d_t + (1 - \beta - \gamma)\mathrm{E}_t[d_{t+1}] + \beta \mathrm{E}_t[p_{t+1}] \qquad (5.4)$$

and

$$d_t = \phi_1 d_{t-1} + \phi_2 d_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}. \qquad (5.5)$$

[1] That is, we require $\frac{1}{2}|\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}| < 1$ and $\frac{1}{2}|\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}| < 1$.

Substituting $\mathrm{E}_t[d_{t+1}] = \phi_1 d_t + \phi_2 d_{t-1} + \theta_1 \varepsilon_t$ into (5.4), we can write p_t as

$$\begin{aligned} p_t = {} & [\gamma + (1-\beta-\gamma)]d_t + \phi_2(1-\beta-\gamma)d_{t-1} \\ & + \theta_1(1-\beta-\gamma)\varepsilon_t + \beta\mathrm{E}_t[p_{t+1}]. \end{aligned} \tag{5.6}$$

Henceforth, we simply refer to equations (5.5) and (5.6) as the *dividend model* and the *pricing equation* without emphasizing the term "log-linearized".

5.2.3 A Rational Expectations Equilibrium

Suppose that the representative investor's *perceived law of motion* (PLM) for p_t is given by

$$p_t = \varphi_1 d_t + \varphi_2 d_{t-1} + \vartheta_1 \varepsilon_t, \tag{5.7}$$

which, as we show later, is consistent with the rational expectations solution of the model; $(\varphi_1, \varphi_2, \vartheta_1)$ are parameters to be determined at the equilibrium. We loosely interpret this investor as a *fundamental investor* because he formulates forecasts of future prices based on fundamental factors like dividends. Equation (5.7) implies that his forecast of p_{t+1}, conditional on his time-t information set, denoted by $\hat{\mathrm{E}}_t[p_{t+1}]$, is given by

$$\hat{\mathrm{E}}_t[p_{t+1}] = (\varphi_2 + \phi_1\varphi_1)d_t + \phi_2\varphi_1 d_{t-1} + \theta_1\varphi_1\varepsilon_t.$$

Substituting $\mathrm{E}_t[p_{t+1}]$ in the pricing equation (5.6) by $\hat{\mathrm{E}}_t[p_{t+1}]$, we obtain the *actual law of motation* (ALM) for p_t,

$$\begin{aligned} p_t = {} & [\gamma + \beta(\varphi_2 + \phi_1\varphi_1) + \phi_1(1-\beta-\gamma)]d_t \\ & + [\beta\phi_2\varphi_1 + \phi_1(1-\beta-\gamma)]d_{t-1} \\ & + [\beta\theta_1\varphi_1 + \theta_1(1-\beta-\gamma)]\varepsilon_t, \end{aligned} \tag{5.8}$$

which shares the same functional form as that of the investor's PLM in (5.7).

Hence, by matching the coefficients of the PLM and the ALM for p_t, we find that, at the *rational expectations equilibrium* (REE), the stock price follows

$$p_t = \overline{\varphi}_1 d_t + \overline{\varphi}_2 d_{t-1} + \overline{\vartheta}_1 \varepsilon_t,$$

where

$$\begin{aligned}
\overline{\varphi}_1 &= 1 + \frac{(\gamma - 1)(\phi_1 + \phi_1\beta - 1)}{\phi_2\beta^2 + \phi_1\beta - 1}, \\
\overline{\varphi}_2 &= \frac{\phi_2(1-\beta)(\gamma - 1)}{\phi_2\beta^2 + \phi_1\beta - 1}, \\
\overline{\vartheta}_1 &= \frac{\theta_1(1-\beta)(\gamma - 1)}{\phi_2\beta^2 + \phi_1\beta - 1}.
\end{aligned} \tag{5.9}$$

5.2.4 A Restricted Perceptions Equilibrium

Now, suppose instead that the representative investor's PLM for p_t is given by

$$p_t = \tau_1 p_{t-1}, \tag{5.10}$$

where τ_1 is a parameter to be determined at the equilibrium. We call this type of investor a *trend follower*. His forecast of p_{t+1}, conditional on his time-t information set, denoted by $\widetilde{\mathrm{E}}_t[p_{t+1}]$, is simply

$$\widetilde{\mathrm{E}}_t[p_{t+1}] = \tau_1 p_t.$$

In Section 5.6, we shall extend above to allow the forecast to depend on some moving average of past stock prices. Our current simple setting nonetheless highlights some key aspects of model formulation and estimation at a later stage.

Substituting $\mathrm{E}_t[p_{t+1}]$ in the pricing equation (5.6) by $\widetilde{\mathrm{E}}_t[p_{t+1}]$, we obtain the ALM for p_t,

$$\begin{aligned} p_t = &\left(\frac{\gamma + \phi_1(1-\beta-\gamma)}{1-\beta\tau_1}\right) d_t + \left(\frac{\phi_2(1-\beta-\gamma)}{1-\beta\tau_1}\right) d_{t-1} \\ &+ \left(\frac{\theta_1(1-\beta-\gamma)}{1-\beta\tau_1}\right) \varepsilon_t, \end{aligned} \tag{5.11}$$

which can be also written in the form

$$p_t = \varphi_1 d_t + \varphi_2 d_{t-1} + \vartheta_1 \varepsilon_t,$$

as in the REE case in Section 5.2.3, for some parameters $(\varphi_1, \varphi_2, \vartheta_1)$ that depend on τ_1. Observe, however, that the investor's PLM for p_t does not coincide with the ALM, so his forecasting model (5.10) is actually *misspecified*. Although the trend follower is using a misspecified model, we require that the forecast errors, namely $p_t - \tau_1 p_{t-1}$, to be orthogonal to the predictor, p_{t-1}. That is, τ_1 satisfies the standard orthogonality condition:

$$\mathrm{E}[(p_t - \tau_1 p_{t-1})p_{t-1}] = 0, \tag{5.12}$$

giving us

$$\tau_1 = (\mathrm{E}[p_{t-1}^2])^{-1}\mathrm{E}[p_t p_{t-1}]. \tag{5.13}$$

To compute the moment equations required in (5.13), it is convenient to express the economic system $(p_t, d_t, d_{t-1}, \varepsilon_t)$ as the following VAR,

$$\begin{bmatrix} p_t \\ d_t \\ d_{t-1} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 0 & \Phi_{12} & \Phi_{13} & \Phi_{14} \\ 0 & \phi_1 & \phi_2 & \theta_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ d_{t-1} \\ d_{t-2} \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \Sigma_1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \varepsilon_t, \tag{5.14}$$

where

$$\Phi_{12} = \frac{\phi_1\gamma + (\phi_2 - \phi_1^2)(1-\beta-\gamma)}{1-\beta\tau_1}, \quad \Phi_{13} = \frac{\phi_2[\gamma + \phi_1(1-\beta-\gamma)]}{1-\beta\tau_1},$$

$$\Phi_{14} = \frac{\theta_1[\gamma + \phi_1(1-\beta-\gamma)]}{1-\beta\tau_1}, \quad \Sigma_1 = \frac{\gamma + (\phi_1+\theta_1)(1-\beta-\gamma)}{1-\beta\tau_1}.$$

More compactly, we can write (5.14) as

$$y_t = \Phi y_{t-1} + \Sigma\varepsilon_t,$$

where $y_t \equiv (p_t, d_t, d_{t-1}, \varepsilon_t)^\top$ and the definitions of Φ and Σ follow directly from the VAR. For y_t to be a covariance-stationary process, we require that all eigenvalues of Φ are within the unit circle. It can be shown that the corresponding eigenvalues are 0, 0, $\frac{1}{2}(\phi_1 - \sqrt{\phi_1^2 + 4\phi_2})$, and $\frac{1}{2}(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})$, and thus y_t is covariance-stationary as long as d_t is so (see Note 1), regardless of τ_1. Hence, assuming stationarity, there exists a 4×4 constant symmetric matrix C_y such that $\mathrm{E}[y_t y_t^\top] = C_y$.

For a VAR(1) model like (5.14), we can explicitly solve for C_y by using

$$(I - \Phi \otimes \Phi)\mathrm{vec}(C_y) = \sigma_\varepsilon^2 \,\mathrm{vec}(\Sigma\Sigma^\top),$$

where I is a conformable identity matrix, $\otimes$ is the Kronecker product operator, and $\mathrm{vec}(\cdot)$ is the vectorization operator. We first solve for $\mathrm{vec}(C_y)$ by solving the linear equations and then obtain $\mathrm{E}[p_t^2]$ ($= \mathrm{E}[p_{t-1}^2]$ by stationarity) as the first diagonal element of C_y. To find $\mathrm{E}[p_t p_{t-1}]$, we compute the matrix

$$\mathrm{E}[y_t y_{t-1}^\top] = \Phi C_y$$

and extract its first diagonal element.

Substituting the solutions of $\mathrm{E}[p_{t-1}^2]$ and $\mathrm{E}[p_t p_{t-1}]$ into (5.13), we can obtain a closed-form solution for τ_1. However, it turns out that the algebraic solution of τ_1, denoted by $\overline{\tau}_1$, is quite lengthy to display, so we do not report it in this article. To illustrate how $\overline{\tau}_1$ is related to the structural parameters (β, γ) nonetheless, we linearly approximate it around $\beta = 1$ and find that

$$\overline{\tau}_1 \approx c_0 + \left(\frac{1-\beta}{\gamma}\right) c_1,$$

where c_0 and c_1 are functions of the dividend model parameters $(\phi_1, \phi_2, \theta_1)$ and are given in Appendix 5.A.

Replacing τ_1 by $\overline{\tau}_1$ in the ALM for p_t in (5.11), we have that, at the *restricted perceptions equilibrium*[2] (RPE), the stock price follows

$$p_t = \overline{\varphi}_1 d_t + \overline{\varphi}_2 d_{t-1} + \overline{\vartheta}_1 \varepsilon_t,$$

where

$$\begin{aligned} \overline{\varphi}_1 &= \frac{\gamma + \phi_1(1-\beta-\gamma)}{1-\beta\overline{\tau}_1}, \\ \overline{\varphi}_2 &= \frac{\phi_2(1-\beta-\gamma)}{1-\beta\overline{\tau}_1}, \\ \overline{\vartheta}_1 &= \frac{\theta_1(1-\beta-\gamma)}{1-\beta\overline{\tau}_1}. \end{aligned} \tag{5.15}$$

[2] We follow the terminology by Evans and Honkapohja (2001). Loosely speaking, we can understand a RPE as an equilibrium between optimally misspecified beliefs and the stochastic processes for the economy.

5.3 Estimating the Model

In this section, we describe how to estimate the dividend model parameters, $(\phi_1, \phi_2, \theta_1, \sigma_\varepsilon^2)$, and the structural parameters, (β, γ).

The model assumes that the dividend and the stock price processes are stationary, which is unlikely to hold in the real world. Instead, suppose that the observed dividends and stock prices are $I(1)$ processes, we can use the following appropriately detrended data for estimation. We first regress the log of dividend on a time trend and an intercept,

$$\log D_t = a_0 + a_1 t + d_t. \tag{5.16}$$

Assuming the residuals d_t is stationary, meaning that $\log D_t$ is trend stationary, we can take d_t as log-dividend deviations. Next we regress the log of stock price on the log of dividend and an intercept,

$$\log P_t = b_0 + b_1 \log D_t + \xi_t. \tag{5.17}$$

Assuming ξ_t is stationary, meaning that $\log P_t$ and $\log D_t$ are cointegrated, substituting (5.16) into (5.17), we can write $\log P_t$ as

$$\log P_t = (b_0 + a_0 b_1) + (a_1 b_1)t + p_t, \tag{5.18}$$

where $p_t \equiv b_1 d_t + \xi_t$ is to be taken as log-price deviations. Note that the long-run price-dividend elasticity is preserved by this detrending procedure, that is, $\partial(\log P_t)/\partial(\log D_t) = \partial p_t/\partial d_t = b_1$.

Given T observations of detrended log-dividends, the ARMA(2,1) dividend model parameters $(\phi_1, \phi_2, \theta_1, \sigma_\varepsilon^2)$ can be estimated by the standard maximum likelihood method. The corresponding log-likelihood is then

$$\mathcal{L}(\phi_1, \phi_2, \theta_1, \sigma_\varepsilon^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma_\varepsilon^2 - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2},$$

where

$$\varepsilon_t = d_t - \phi_1 d_{t-1} - \phi_2 d_{t-2} - \theta_1 \varepsilon_{t-1},$$

for $t = 3, \ldots, T$ and we set $\varepsilon_1 = \varepsilon_2 = 0$.

To obtain an estimate of β, we use the equilibrium interest rate equation (5.2) and observe that

$$\beta = \frac{1}{1+r}.$$

Since in the real world the interest rate is nonconstant, we need to determine an appropriate r. Given data of interest rates, say $\{r_t\}$, we regress r_t on a constant only. We denote the estimate by $\hat{r}$ and the associated Newey-West standard error by $\text{Se}(\hat{r})$. We obtain the estimate $\hat{\beta}$ by replacing r by $\hat{r}$. Next, applying the delta method, the standard error of $\hat{\beta}$ can be approximated by using

$$\text{Se}(\hat{\beta}) \approx \frac{\text{Se}(\hat{r})}{(1+\hat{r})^2}.$$

The only remaining unknown parameter to estimate is γ and we propose to estimate it using the generalized method of moments (GMM) approach. The basic idea of our estimation strategy is to use the model PLM and the ALM for p_t to generate a set of orthogonality conditions, $\vec{g}_T$, for the GMM estimation. Specifically, given T observations of $(p_t, d_t, \varepsilon_t)$, we define $\vec{g}_T$ by

$$\vec{g}_T(\gamma) \equiv \begin{cases} \frac{1}{T}\sum_{t=1}^{T}\{[p_t - (\overline{\varphi}_1 d_t + \overline{\varphi}_2 d_{t-1} + \overline{\vartheta}_1 \varepsilon_t)] \otimes z_t\}, & \text{under REE,} \\ \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\{[p_t - (\overline{\varphi}_1 d_t + \overline{\varphi}_2 d_{t-1} + \overline{\vartheta}_1 \varepsilon_t)] \otimes z_t\} \\ \frac{1}{T}\sum_{t=1}^{T}\{(p_t - \overline{\tau}_1 p_{t-1})p_{t-1}\} \end{bmatrix}, & \text{under RPE,} \end{cases}$$

where $z_t \equiv (d_t, d_{t-1}, \varepsilon_t)^\top$ and we set $p_0 = d_0 = 0$. We note that there is one more moment condition under the RPE because the PLM for p_t is different from the ALM for p_t.

Our GMM estimate $\hat{\gamma}$ is taken as the limit of a sequence of $\hat{\gamma}_{(i)}$ that solves the following iterative scheme:

$$\hat{\gamma}_{(i+1)} = \arg\min_{\gamma>0} J_T(\gamma) \equiv T\, \vec{g}_T(\gamma)^\top \vec{S}_T^{-1}\, \vec{g}_T(\gamma),$$

where $\vec{S}_T$ is the Newey-West estimate given by

$$\vec{S}_T = \vec{\Gamma}_{0,T} + \sum_{v=1}^{q}\{[1 - v/(q+1)](\vec{\Gamma}_{v,T} + \vec{\Gamma}_{v,T}^\top)\},$$

$$\vec{\Gamma}_{v,T} = \frac{1}{T}\sum_{t=v+1}^{T} \vec{g}_T(\hat{\gamma}_{(i)})\vec{g}_T(\hat{\gamma}_{(i)})^\top.$$

We follow the Newey-West suggestion to set q to be the nearest integer of $4(T/100)^{2/9}$. The resulted $\hat{\gamma}$ is commonly referred to as an iterated GMM estimate. The standard error of $\hat{\gamma}$ is given by

$$\mathrm{Se}(\hat{\gamma}) = \sqrt{\vec{V}_T/T},$$

where

$$\vec{V}_T = \{\vec{D}_T \vec{S}_T^{-1} \vec{D}_T^\top\}^{-1} \quad \text{and} \quad \vec{D}_T = \frac{\partial \vec{g}_T(\hat{\gamma})}{\partial \gamma}.$$

Note that in the GMM estimation we essentially take the estimated values of $(\phi_1, \phi_2, \theta_1, \sigma_\varepsilon^2, \beta)$ as true values. We provide two justifications for this treatment. First, it turns out that the standard errors of these estimates are small relative to that of $\hat{\gamma}$, so we expect little distortion to $\text{Se}(\hat{\gamma})$ even if they are taken into account. Second, this treatment allows us to more reliably estimate γ because it is well-known that numerical optimization algorithms are less accurate for higher-dimension estimation problems.

5.4 Testing the Model

5.4.1 Testing the Overidentifying Restrictions

Suppose that $\vec{g}_T$ contains ℓ moment conditions, then under the null hypothesis that the model is "valid", the test statistic

$$J_T(\hat{\gamma}) = T\, \vec{g}_T(\hat{\gamma})^\top \vec{S}_T^{-1}\, \vec{g}_T(\hat{\gamma}),$$

where $\hat{\gamma}$ is the iterated GMM estimate, is asymptotically Chi-squared distributed with $\ell - 1$ degrees of freedom. The alternative hypothesis is that the model is "invalid". We reject the null hypothesis at the $\mathfrak{a}$-percent significance level if $J_T(\hat{\gamma})$ is greater than the $(1 - \mathfrak{a})$-th percentile of the $\chi^2_{\ell-1}$ distribution. A rejection of the null hypothesis is taken as evidence against the model. This statistical test is commonly referred to as Hansen's χ^2 test and the quantity J_T/T is called the Hansen and Jannathan (1991) distance, or simply the HJ distance.

5.4.2 Comparing Model Explanatory Powers under the REE and the RPE

Given T observations of $(p_t, d_t, \varepsilon_t)$, denote the model equilibrium price by $\hat{p}_t = \overline{\varphi}_1 d_t + \overline{\varphi}_2 d_{t-1} + \overline{\vartheta}_1 \varepsilon_t$, where we recall that the definitions of $(\overline{\varphi}_1, \overline{\varphi}_2, \overline{\vartheta}_1)$ under the REE and the RPE are given in (5.9) and (5.15). We can view $\hat{p}_t$ as a prediction of the observed p_t. We calculate the mean squared error (MSE) of the predictor $\hat{p}_t$ by

$$\mathrm{MSE}(\hat{\gamma}) = \frac{1}{T}\sum_{t=1}^{T}(p_t - \hat{p}_t)^2,$$

which is a measure of the differences between log-price deviations predicted by the model and the values actually observed. The function argument emphasizes that the MSE depends on the GMM estimate $\hat{\gamma}$.

Let $j = 1, 2$ index the model equilibria under the REE and the RPE, respectively. Let $\mathrm{E}[\vec{g}_T(\gamma_j)]$ denote the vector of population average of the moment conditions, that is, the expectation is taken with respect to the true joint density of $\{(p_t, d_t, \varepsilon_t)\}_{t=1}^{T}$. Let also

$$\gamma_j^* \equiv \arg\min_{\gamma_j > 0} \mathrm{E}[\vec{g}_T(\gamma_j)]^\top \{\mathrm{E}[\vec{S}_T]\}^{-1} \mathrm{E}[\vec{g}_T(\gamma_j)]$$

and $f^* \equiv \mathrm{MSE}(\gamma_1^*) - \mathrm{MSE}(\gamma_2^*)$. The null hypothesis is

$$\mathrm{H}_0 : \mathrm{E}[f^*] \leq 0,$$

which states that the equilibrium pricing model under the REE has superior explanatory power over that of the RPE.

This hypothesis can be tested based on the “reality check” method of White

(2000). The general testing strategy is to generate bootstrap samples, denoted by $\{(p_t^{(k)}, d_t^{(k)}, \varepsilon_t^{(k)})\}_{t=1}^T$, $k = 1, \ldots, K$, and for each bootstrap sample we compute $\hat{f}_{(k)} \equiv \text{MSE}(\hat{\gamma}_1^{(k)}) - \text{MSE}(\hat{\gamma}_2^{(k)})$, where $\hat{\gamma}_1^{(k)}$ and $\hat{\gamma}_2^{(k)}$ are the associated iterated GMM estimates. To apply White's reality check test, we employ the test statistic given by

$$\hat{f} \equiv \text{MSE}(\hat{\gamma}_1) - \text{MSE}(\hat{\gamma}_2)$$

and compute the bootstrap estimate of the p-value given by

$$\mathfrak{p} \equiv \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}\{\hat{f}_{(k)} > \hat{f}\},$$

where the notation $\mathbf{1}$ represents an indicator function. Since this is a one-side test, at the $\mathfrak{a}$-percent significance level, the critical value is the value at the $(1-\mathfrak{a})$-th percentile of the bootstrap test statistics $\hat{f}_{(k)}$. We do not reject the null if the test statistic $\hat{f}$ is not unusually high, *i.e.*, is not greater than the value at the $(1-\mathfrak{a})$-th percentile. Therefore, the test rejects the null hypothesis if $\mathfrak{p}$ is less than $\mathfrak{a}$ percent. We note that Chen and Ludvigson (2009) propose a similar statistical methodology to compare the HJ distances of competing asset pricing models. We instead compare MSEs because MSE is a more common and intuitive measure of model explanatory power. However, both measures should give us consistent quantitive conclusions.

We base our bootstrap resampling on an estimated VAR(q) model for (p_t, d_t),

$$\begin{bmatrix} p_t \\ d_t \end{bmatrix} = \Phi_1 \begin{bmatrix} p_{t-1} \\ d_{t-1} \end{bmatrix} + \Phi_2 \begin{bmatrix} p_{t-2} \\ d_{t-2} \end{bmatrix} + \cdots + \Phi_q \begin{bmatrix} p_{t-q} \\ d_{t-q} \end{bmatrix} + \begin{bmatrix} e_{pt} \\ e_{dt} \end{bmatrix},$$

where the lag q is selected by the Akaike information criterion (AIC). Each bootstrap sample is generated by resampling from the residual pairs, (e_{pt}, e_{dt}), with replacement, and then solving the model recursively. To better capture structural changes in the residuals, we perform the multiple breakpoint test by Bai and Perron (2003) on the squared residuals, (e_{pt}^2, e_{dt}^2), and allow residual resampling distributions to differ across the estimated break dates. We present the VAR estimation results and the Bai-Perron test results in Appendix 5.B.

5.5 Empirical Results

5.5.1 Obtaining Detrended Data and Relevant Tests

We obtain time-series data for real interest rate (r_t), real S&P 500 index price (P_t), and real S&P 500 dividend (D_t). The real interest rate is taken as the difference of the US 30-day Treasury yield and the growth rate of the US CPI (for all urban consumers). The inflation adjustment to the S&P500 price and dividend data is also based on the historic US CPI. The sample begins as early as 1871:01 for the index price, the dividend, and the CPI data to as late as 1926:01 for the Treasury yield data and ends in 2015:12 for all of the series.

Estimating equation (5.16), we obtain the following:

$$\underset{\text{(Se)}}{\log D_t} = \underset{(0.023608)}{1.728845} + \underset{(2.10\times10^{-5})}{0.000899t} + d_t,$$

$$\text{Adjusted } R^2 = 0.850937, \quad \text{MSE} = 0.035734,$$

where $t = 1, 2, \ldots, 1740$ is time measured chronologically. The reported standard errors are Newey-West (NW) standard errors. The t-statistic for the deterministic trend is 42.915, which appears to be strong evidence for a linear trend. However, it is also important to check whether the residuals d_t from this model appear to be stationary.

We then use the augmented Dickey-Fuller (ADF) test (without an intercept and a trend because d_t are residuals) to test for the presence of a unit root in d_t. The lag length is optimally selected by the AIC. We find that the test statistic is -4.299 and the 1-percent level critical value is -2.566.[3] Therefore, we strongly reject the null hypothesis that d_t has a unit root.

We note that one could also use the ADF test (with an intercept and a linear trend) to test for the presence of a unit root in $\log D_t$. In that case, the test statistic would be -4.285 and the 1-percent level critical value would be -3.963. In both cases, the conclusion that real dividend appears to be trend stationary holds.[4] Nonetheless, we prefer directly testing d_t because our later model estimation is based on this series.

Before we estimate equation (5.17) to show that $\log P_t$ and $\log D_t$ are cointegrated, we first illustrate that they are both $I(1)$ processes. We use the ADF test (with an intercept but no trend) to test for the presence of a unit root in the first difference of $\log P_t$ and $\log D_t$ one at a time. The lag lengths are again optimally selected by the AIC. We find that the test statistic is -10.298 for the first difference of $\log P_t$, and -8.563 for the first difference of $\log D_t$. The corresponding 1-percent critical value for these tests is -3.434. Therefore, there is strong evidence that $\log P_t$ and $\log D_t$ appear to be $I(1)$ processes.

[3] Although the 5-percent level is more commonly used in the literature, this choice is arbitrary. Since we want to be particularly confident in our results, we set a more stringent level of 1 percent throughout this article.

[4] These two approaches are actually not equivalent: when d_t is used, one must assume that the slope of the linear trend is accurately estimated.

Estimating equation (5.17), we obtain the following:

$$\underset{(\text{Se})}{\log P_t} = \underset{(0.099065)}{1.672064} + \underset{(0.04051)}{1.611302} \log D_t + \xi_t,$$

$$\text{Adjusted } R^2 = 0.867738, \quad \text{MSE} = 0.094858,$$

where we use the NW standard errors. The t-statistic for the slope coefficient, b_1, is 39.834, which appears to be strong evidence for a long-run linear relationship between these two variables. However, it is also important to check whether the residuals ξ_t from this model appear to be stationary (if so, $\log P_t$ and $\log D_t$ are said to be cointegrated), otherwise we have a spurious regression.

We follow the Engle-Granger (EG) methodology to test for cointegration. First, we use the ADF test (with no intercept and trend) to test for the presence of a unit root in ξ_t. We find that the test statistic is -4.037 and the 1-percent level critical value is -2.566. Therefore, we strongly reject the null hypothesis that ξ_t has a unit root. Next, the EG methdology supplements the ADF test result with the following error-correction model. If $\log P_t$ and $\log D_t$ are cointegrated, these variables admit the error-correction form:

$$\Delta \log P_t = \alpha_1 + \alpha_p \xi_{t-1} + \sum_{j=1}^{q} \alpha_{11}(j) \Delta \log P_{t-j} + \sum_{j=1}^{q} \alpha_{12}(j) \Delta \log D_{t-j} + \varepsilon_{pt},$$

$$\Delta \log D_t = \alpha_2 + \alpha_d \xi_{t-1} + \sum_{j=1}^{q} \alpha_{21}(j) \Delta \log P_{t-j} + \sum_{j=1}^{q} \alpha_{22}(j) \Delta \log D_{t-j} + \varepsilon_{dt},$$

where the lag length q is to be optimally selected by the AIC. We know that at least α_p or α_d (often called *speeds of adjustment*) should be significantly different from zero if the variables are cointegrated. Also, direct convergence necessitates that α_p be nonpositive and α_d be nonnegative. Furthermore, their absolute values should not be too large as we are modelling how $\Delta \log P_t$ and $\Delta \log D_t$ converge to their *long-run* equilibrium relationship.

Our results show that $q = 6$ is optimal by the AIC. The estimate of α_p is -0.004850 with the standard error being 0.00327 and the t-statistic being -1.48203. The estimate of α_d is 0.006888 with the standard error being 0.00126 and the t-statistic being 5.4862. We find that α_d is significantly different from zero at the 1-percent level (although α_p is not). Observe also that both the magnitude and the signs of α_p and α_d are as expected. Therefore, based on the EG methodology, we find strong evidence that $\log P_t$ and $\log D_t$ are cointegrated.

The significance of the above results is that we can justify why it is "appropriate" to detrend $\log P_t$ using equation (5.18) to obtain p_t. In particular, we realize that if one used the ADF test (with an intercept and a linear trend) to test whether $\log P_t$ has a unit root, the test statistic would be -2.696 while the 1-percent level critical value would be -3.963 (the 5 and 10-percent critical values would be -3.412 and -3.128). Thus, one might conclude that $\log P_t$ is not trend stationary and so the usual linear detrending would be considered "inappropriate" (because the residuals would not be stationary either). It is beyond the intended scope of this article to further investigate whether $\log P_t$ is trend stationary (indeed it is possible that different testing procedures give different conclusions), we only aim to illustrate that when $\log P_t$ is taken as the log of inflation adjusted S&P 500 index price, the trend stationarity assumption is not unrealistic.

5.5.2 Estimating the Dividend Model Parameters

Using the detrended dividend data d_t, we estimate the dividend model given by (5.5) and obtain the following:

$$\underset{\text{(Se)}}{d_t} = \underset{(0.021974)}{1.879657}d_{t-1} - \underset{(0.021882)}{0.882386}d_{t-2} + \varepsilon_t - \underset{(0.034276)}{0.676361}\varepsilon_{t-1},$$

$$\text{Adjusted } R^2 = 0.993457, \quad \sigma_\varepsilon^2 = (0.015299)^2, \quad \text{Log likelihood} = 4800.181.$$

We find that all parameters are significantly different from zero at the 1-percent level. We calculate that the AR roots are 0.969686 and 0.909970, which are both within the unit circle, this indicates that this ARMA model is stationary (see Note 1). Although not explicitly used in this article, we calculate that the MA root is 0.67661, this indicates that this ARMA model is also invertible.

We then proceed to check serial correlation in the residuals ε_t based on the Ljung-Box Q-statistics for lags up to 1 to 12. The null hypothesis is that the residuals do not exhibit serial correlation (up to the specified number of lags). We cannot reject the null hypothesis for lags up 4 at the 1-percent level. However, as the lag number increases, we consistently reject the null hypothesis at the 1-percent level. Although there is some evidence that the residuals are not completely uncorrelated, they are sufficiently uncorrelated in shorter horizons, implying that not much "recent" information is wasted for constructing predictions of d_t (*i.e.*, given the model specification, no recent information can further improve the predictions).

We try other simpler ARMA specifications such as $(1, 0)$, $(2, 0)$, and $(1, 1)$ but find that the residuals, by Q-statistics, are correlated even at lower lags (with PAC $>$ 0.1). We also estimate an ARMA(3,1) model and an ARMA(2,2) model but find that the AR(3) and the MA(2) coefficients are quite insignificant (with p-values of 0.1905 and 0.1309, respectively). Therefore, we conclude that an ARMA(2,1) model appears to be adequate to approximate the actual dynamics of d_t.

5.5.3 Estimating the Structural Parameters

Regressing r_t on a linear time trend and an intercept, we find that the trend is highly insignificant (the estimate is 6.20×10^{-9} with the NW standard error being 1.06×10^{-06} and the p-value being 0.9953). Thus, we conclude that it is not necessary to detrend the observed real interest rate data. Next, regressing r_t on an intercept only, we obtain the intercept estimate $\hat{r} = 0.000418$ with the NW standard error $\mathrm{Se}(\hat{r}) = 0.000283$. This gives us the estimate of β and its standard error (by delta method) as follows:

$$\hat{\beta} = \frac{1}{1+\hat{r}} = 0.999583 \quad \text{and} \quad \mathrm{Se}(\hat{\beta}) \approx \frac{\mathrm{Se}(\hat{r})}{(1+\hat{r})^2} = 2.83 \times 10^{-4}.$$

The 99-percent asymptotic confidence interval is $[0.998930, 1.000235]$, with the understanding that the upper bound should be taken as 1.

Substituting the estimates (taken as their true values)

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \theta_1 \\ \sigma_\varepsilon^2 \\ \beta \end{bmatrix} = \begin{bmatrix} 1.879657 \\ -0.882386 \\ -0.676361 \\ (0.015299)^2 \\ 0.999583 \end{bmatrix}$$

into the vector of orthogonality conditions $\vec{g}_T$, we can now estimate γ by the GMM approach. Table 5.1 displays the estimation results.

Table 5.1. GMM Estimation Results under the REE and the RPE

	$\hat{\gamma}$	$\mathrm{Se}(\hat{\gamma})$	J_T	DF	p-value
REE	1.427271	0.345553	6.867459	2	0.032266
RPE	0.002069	0.009244	56.026102	3	4.15×10^{-12}

We obtain economically plausible estimates of γ under both the REE and the RPE. Under the assumption of constant relative risk averse preferences, our results imply that the representative investor has to be more risk tolerant under the RPE such that the model can possibly explain the observed stock prices (we discuss model testing in the next section). This finding is consistent with the intuition that trend followers should be more risk tolerant than fundamental investors. Indeed, under the RPE, the t-statistic for the null hypothesis that $\gamma = 0$ is 0.2238 while the t-statistic is 4.1304 under the REE. Therefore, at the 1-percent significance level, we cannot reject the null hypothesis that the representative investor is risk neutral under the RPE but we reject the null hypothesis under the REE.

5.6 Testing the Model

5.6.1 Testing the Overidentifying Restrictions

Table 5.1 also presents the Hansen's χ^2 tests. We know from Section 5.4.1 that under the null hypothesis that the model is "valid", the J_T-statistic is asymptotically Chi-squared distributed with degrees of freedom (DF) being the number of overidentifying restrictions. We reject the null hypothesis at the $\mathfrak{a}$-percent significance level if J_T is greater than the $(1-\mathfrak{a})$-th percentile of the χ^2_{DF} distribution. Equivalently, we reject the null hypothesis if the associated p-value is less than $\mathfrak{a}$ percent. Our results are as follows.

We do *not* reject the model under the REE at the 1-percent significance level (although it would be rejected at the 5-percent level). This result does imply that assuming all investors aggregately behave as if they were a single fundamental investor is not as implausible as it sounds. By contrast, the model under the REP is strongly rejected at all conventional significance levels. This result implies that assuming all investors aggregately behave as if they were a single trend follower does not admit a model that can possibly generate our actual data.

We further re-estimate the model under the RPE using only the first four orthogonality conditions in $\vec{g}_T$, *i.e.*, ignoring the orthogonality condition implied by the PLM for p_t. In this case, we have $\hat{\gamma} = 0.002075$, $\mathrm{Se}(\hat{\gamma}) = 0.009244$, $J_T = 55.362250$, DF $= 2$, and p-value $= 9.51 \times 10^{-13}$. We see that the results are largely similar to that of the "full" estimation. Hence, our previous rejection of the model is *not* due to an additional orthogonality condition compared to the REE case. This finding allows us to specifically conclude that the model under the RPE does not admit an equilibrium pricing equation (or ALM) that can well describe the observed price dynamics.

5.6.2 Comparing Model Explanatory Powers under the REE and the RPE

Examining the numerical results presented in Table 5.2, we see that the reality check p-value is 0.258, which is higher than all conventional significance levels. Therefore, we do *not* reject the null hypothesis that the equilibrium pricing model under the REE has superior explanatory power over that of the RPE. This result is taken as assurance that the model under the REE can genuinely explain better the observed price dynamics, unlikely just by luck.

To summarize, we have evidence that the model under the REE has superior econometric performance compared to that of the RPE. This may not be surprising, but without formulating a model and testing it with data, there would be no way to tell that the RPE, while theoretically possible, is empirically poorly supported.

Table 5.2. Reality Check Results: Mean Squared Error Performance

	REE	RPE
MSE, $\frac{1}{T}\sum(p_t - \hat{p}_t)^2$	0.093232	0.143040
Difference in MSE, $\hat{f}$	-0.049809	
Reality check p-value, $\mathfrak{p}$	0.258	
Number of bootstrap samples	1,000	

Notes: The null hypothesis is that the equilibrium pricing model under the REE has superior explanatory power over that of the RPE. This hypothesis is tested based on the "reality check" method by White (2000).

5.7 Extension: Technical Forecasting with Moving Averages

Let us consider an exponentially weighted *moving average* (MA) defined by

$$a_t \equiv \alpha \sum_{j=-\infty}^{t} (1-\alpha)^{t-j} p_{j-1},$$

where $\alpha \in (0,1)$ is a given parameter controlling the window size. We note that this definition implies that

$$a_{t+1} = (1-\alpha)a_t + \alpha p_t.$$

Since a_{t+1} is a weighted average of a_t and p_t, this quantity is known given the time-t information set. In practice, one set

$$\alpha = \frac{2}{N+1}$$

and call the corresponding a_t the N-period moving average. Thus, the lag p_{t-1} is equivalently to the 1-period MA.

Now, let us modify the representative investor's PLM for p_t, given by equation (5.10) in Section 5.2.4, by replacing p_{t-1} by a_t. That is, his PLM is given by

$$p_t = \tau_1 a_t,$$

where τ_1 is a parameter to be determined at the equilibrium. We call this type of investor a *technical investor*. His forecast of p_{t+1}, conditional on his time-t information set, denoted by $\widetilde{\mathrm{E}}_t[p_{t+1}]$, becomes

$$\widetilde{\mathrm{E}}_t[p_{t+1}] = \tau_1 a_{t+1} = \tau_1(1-\alpha)a_t + \tau_1 \alpha p_t.$$

The corresponding ALM for p_t can be written as

$$p_t = \varphi_1 d_t + \varphi_2 d_{t-1} + \vartheta_1 \varepsilon_t + \lambda_1 a_t,$$

where

$$\begin{aligned} \varphi_1 &= \frac{\gamma + \phi_1(1-\beta-\gamma)}{1-\alpha\beta\tau_1}, \quad \varphi_2 = \frac{\phi_2(1-\beta-\gamma)}{1-\alpha\beta\tau_1}, \\ \vartheta_1 &= \frac{\theta_1(1-\beta-\gamma)}{1-\alpha\beta\tau_1}, \quad \lambda_1 = \frac{\beta\tau_1(1-\alpha)}{1-\alpha\beta\tau_1}. \end{aligned} \tag{5.19}$$

Following Section 5.2.4, we require τ_1 to satisfy the standard orthogonality condition:

$$\mathrm{E}[(p_t - \tau_1 a_t)a_t] = 0,$$

giving us

$$\tau_1 = (\mathrm{E}[a_t^2])^{-1}\mathrm{E}[p_t a_t]. \tag{5.20}$$

To compute the moment equations required in (5.20), we can make use of the the following VAR,

$$\begin{bmatrix} p_t \\ d_t \\ d_{t-1} \\ \varepsilon_t \\ a_t \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} & \Phi_{14} & \Phi_{15} \\ 0 & \phi_1 & \phi_2 & \theta_1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 & 1-\alpha \end{bmatrix} \begin{bmatrix} p_{t-1} \\ d_{t-1} \\ d_{t-2} \\ \varepsilon_{t-1} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} \Sigma_1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \varepsilon_t,$$

where

$$\Phi_{11} = \frac{\alpha(1-\alpha)\beta\tau_1}{1-\alpha\beta\tau_1}, \quad \Phi_{12} = \frac{\phi_1\gamma + (\phi_2 - \phi_1^2)(1-\beta-\gamma)}{1-\alpha\beta\tau_1},$$

$$\Phi_{13} = \frac{\phi_2[\gamma + \phi_1(1-\beta-\gamma)]}{1-\alpha\beta\tau_1}, \quad \Phi_{14} = \frac{\theta_1[\gamma + \phi_1(1-\beta-\gamma)]}{1-\alpha\beta\tau_1},$$

$$\Phi_{15} = \frac{(1-\alpha)^2\beta\tau_1}{1-\alpha\beta\tau_1}, \quad \Sigma_1 = \frac{\gamma + (\phi_1+\theta_1)(1-\beta-\gamma)}{1-\alpha\beta\tau_1}.$$

Given the functions of $\mathrm{E}[a_t^2]$ and $\mathrm{E}[p_t a_t]$, which are in terms of τ_1, by equation (5.20), the solution for τ_1, denoted by $\overline{\tau}_1$, solves the function h given by

$$h(\tau_1) \equiv (\mathrm{E}[a^2(\tau_1)])^{-1}\mathrm{E}[p(\tau_1)a(\tau_1)] - \tau_1 = 0. \tag{5.21}$$

It turns out that $\overline{\tau}_1$ does not admit a closed-form solution. However, given some appropriate quantity τ_1^0, we can linearly approximate h by

$$h(\tau_1) \approx h(\tau_1^0) + \frac{\partial h(\tau_1^0)}{\partial \tau_1}(\tau_1 - \tau_1^0), \tag{5.22}$$

implying an approximate solution

$$\overline{\tau}_1 \approx \hat{\tau}_1 \equiv \tau_1^0 - h(\tau_1^0)\left(\frac{\partial h(\tau_1^0)}{\partial \tau_1}\right)^{-1}.$$

To obtain a reasonable τ_1^0, we can use data to regress p_t on a_t (with no intercept) and take the least-squares estimate as τ_1^0.

Substituting τ_1 by $\hat{\tau}_1$ in (5.19) and denoting the corresponding parameters by $(\hat{\varphi}_1, \hat{\varphi}_2, \hat{\vartheta}_1, \hat{\lambda}_1)$, we can estimate γ by using the following set of orthogonality conditions by the GMM approach as discussed in Section 5.3 with

$$\vec{g}_T(\gamma) \equiv \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\{[p_t - (\hat{\varphi}_1 d_t + \hat{\varphi}_2 d_{t-1} + \hat{\vartheta}_1 \varepsilon_t + \hat{\lambda}_1 a_t)] \otimes z_t\} \\ \frac{1}{T}\sum_{t=1}^{T}\{(p_t - \hat{\tau}_1 a_t)a_t\} \end{bmatrix},$$

where $z_t \equiv (d_t, d_{t-1}, \varepsilon_t, a_t)^\top$. We then use the Hansen's χ^2 test to test the model validity as described in Section 5.4.1. However, because we approximate the actual solution $\overline{\tau}_1$ by $\hat{\tau}_1$, it is possible that a rejection of the model is due to a poor approximation. Therefore, we propose to supplement the Hansen's χ^2 test with the following Wald test.

Since $\hat{\tau}_1$ is a function of γ, substituting it into h given by (5.21), we can view h as a function of γ. Define $\hat{h}(\gamma) \equiv h(\hat{\tau}_1)$. We test the null hypothesis that

$$\mathrm{H}_0 : \hat{h}(\gamma) = 0$$

by using the Wald test statistic given by

$$W_T(\hat{\gamma}) \equiv [\hat{h}(\hat{\gamma})]^2 \left\{ \mathrm{Se}(\hat{\gamma}) \times \frac{\partial \hat{h}(\hat{\gamma})}{\partial \gamma} \right\}^{-2},$$

where $\hat{\gamma}$ is the iterated GMM estimate and $\mathrm{Se}(\hat{\gamma})$ is its standard error. We reject the null hypothesis at the $\mathfrak{a}$-percent significance level if $W_T(\hat{\gamma})$ is greater than the $(1-\mathfrak{a})$-th percentile of the χ_1^2 distribution. Failing to reject the null hypothesis is taken as some assurance that the approximation is reasonably accurate. That is, at least the function $\hat{h}$ evaluated at $\hat{\gamma}$ is not "statistically" far from zero.

In practice, moving averages are commonly measured in days. Since we use monthly data, assuming there are 30 days a month, we approximate the L-day MA by the N-month MA by setting $N = \lfloor L/30 \rceil$, where "$\lfloor \cdot \rceil$" represents the nearest integer function. In this article, we consider the popular 50, 100, and 200-day MAs. Thus, recalling $\alpha = 2/(N+1)$, we set α to be 0.6667, 0.5, and 0.25, respectively.

Table 5.3 displays the GMM estimation results under each of these three cases. Similar to our previous RPE described in Section 5.2.4, the model requires the representative investor, now labelled as a technical investor, to be almost risk neutral such that it can possibly explain the observed stock prices. In fact, at the 1-percent significance level, we cannot reject the null hypothesis that the representative investor is risk neutral under all three cases. Table 5.3 also presents the Hansen's χ^2 tests. The tests once again reveal that at least one of the sample moment restrictions is violated. Thus, we conclude that the observed stock prices are *not* consistent with any of these three RPEs. Furthermore, we see from Table 5.4 that the p-values associated with the Wald tests are quite large, so we cannot reject the null hypothesis that $\hat{h}(\gamma) \equiv h(\hat{\tau}_1) = 0$. This indicates that the rejection of the model by the Hansen's χ^2 tests is not due to a poor linear approximation of $h(\tau_1)$ given by (5.22).

Table 5.3. GMM Estimation Results under the Other Three RPEs

	$\hat{\gamma}$	Se($\hat{\gamma}$)	J_T-statistic	DF	p-value
RPE (50-day MA)	6.08×10^{-21}	0.057812	79.582968	4	2.13×10^{-16}
RPE (100-day MA)	4.89×10^{-20}	0.122686	79.079805	4	2.73×10^{-16}
RPE (200-day MA)	6.76×10^{-16}	0.230710	79.644104	4	2.07×10^{-16}

Notes: The 50, 100, and 200-day MAs are approximated by the 2, 3, and 7-month MAs of monthly stock prices. The translation is based on the fact that the width of the L-day MA window approximately equals that of the N-month MA, where N is the nearest integer of $L/30$.

Table 5.4. Wald Test Results under the Other Three RPEs

	W_T-statistic	DF	p-value
RPE (50-day MA)	0.021814	1	0.882584
RPE (100-day MA)	0.005115	1	0.942985
RPE (200-day MA)	0.001553	1	0.968561

Notes: The null hypothesis is that $\hat{h}(\gamma) \equiv h(\hat{\tau}_1) = 0$. Failing to reject the null hypothesis is taken as assurance that the linear approximation of $h(\tau_1)$ given by (5.22) is reasonably accurate.

5.8 Discussion

While the results of estimation appears to be discouraging for the use of technical analysis, this is not necessarily true. The reason is, as we learnt from Chapter 4, that even if prices are fully "rational", technical indicators such as moving averages are still useful in forecasting price trends because the stock prices are strongly first-order correlated. However, one may question why an individual would forecast prices based on technical analysis but not dividends directly. A likely answer is that dividends may not be observed at a desired frequency useful for the investors. Not to mention aggregate dividends are difficult to calculate. Therefore, investors need to seek an alternative forecasting strategy.

Technical analysis therefore provides a simple solution as it only requires easily obtainable data such as past prices. Hence, it is reasonable to imagine that investors can rely on technical analysis for higher frequency trading but once fundamental information is released, typically at a much lower frequency compared to stock prices, they adjusts their expectations of prices and thus prices reflect such fundamental changes accordingly. Since our estimation is based on monthly data, we have no direct evidence against the use of technical analysis for high frequency trading.

5.9 Conclusion

In this article, adapting Lucas' (1978) one-tree representative investor framework, we have formulated a model to study the predictive power of moving averages as an equilibrium phenomenon. A special feature of this model is that prices and the parameters of the investor's stock price forecasting model are all determined endogenously, as functions of structural parameters representing time preference and risk aversion. By imposing different assumptions on the investor's forecasting model, we obtain several possible equilibria. The equilibrium pricing equation and the investor's forecasting model imply a set of orthogonality conditions which provides the basis for estimation and testing strategies to examine whether the proposed equilibria are empirically supported.

The equilibria considered in this article are labelled as either a rational expectations equilibrium (REE) or a restricted perceptions equilibrium (RPE). A REE prevails when the investor's forecasting model coincides with the equilibrium pricing equation. We show that this is the case if the investor formulates forecasts of future prices based on dividends and observable shocks to dividends. A RPE prevails when the investor misspecifies his forecasting model but his forecasts are otherwise optimal within a limited class of forecasting models. We show that it is possible to arrive at a RPE when the investor ignores dividends but forecasts stock prices purely based on lagged prices or moving averages of past prices.

Using actual monthly data on dividends and stock index prices, we however show that while the RPEs are theoretically attainable, they are empirically poorly supported. By contrast, we find that the REE is more plausible with superior econometric performance.

5.10 Appendix

Appendix 5.A. Definitions of c_0 and c_1 in Section 5.2.4

In Section 5.2.4, we state that

$$\overline{r}_1 \approx c_0 + \left(\frac{1-\beta}{\gamma}\right) c_1,$$

when β is close to one. The coefficients c_0 and c_1 are defined as follows:

$$c_0 \equiv \frac{\phi_1^2(\theta_1 - \phi_2 - 1) + (1+\phi_2)(\phi_2 - \theta_1)(\theta_1 - 1)}{1 + \phi_2^2 - \phi_1(1+\phi_2 - 2\theta_1) - 2\phi_1\theta_1 + 2\theta_1(\theta_1 - 1)} + \frac{\phi_1[\phi_2 + \phi_1^2 + (\theta_1 - 1)^2 - 2\phi_2\theta_1]}{1 + \phi_2^2 - \phi_1(1+\phi_2 - 2\theta_1) - 2\phi_1\theta_1 + 2\theta_1(\theta_1 - 1)},$$

and

$$c_1 \equiv \frac{(1+\phi_1-\phi_2)(1+\phi_2)(\phi_1 - \phi_2 + 2\theta_1 - 1)[\theta_1(\phi_1+\theta_1) - \phi_2]}{[1 + \phi_2^2 - \phi_1(1+\phi_2 - 2\theta_1) - 2\phi_1\theta_1 + 2\theta_1(\theta_1 - 1)]^2}.$$

Their values are determined once we estimate the dividend model parameters $(\phi_1, \phi_2, \theta_1)$. Our estimate of β is 0.999583, so the approximate formula above should be reasonably accurate. However, we recall that the purpose of this approximation is to more conveniently present how $\overline{\tau}_1$ is related to the structural parameters (β, γ). The actual closed-form solution of $\overline{\tau}_1$ is used in our estimation.

Appendix 5.B. Estimated VAR Model for Bootstrap Reality Check

Using detrended S&P 500 dividend and index price monthly data for the period 1871:1 to 2015:12, we obtain the following estimated VAR model for bootstrap reality check described in Section 5.4.2:

$$\begin{aligned}
p_t = {} & 1.276497p_{t-1} - 0.336224p_{t-2} + 0.022018p_{t-3} + 0.070916p_{t-4} \\
& + 0.023310p_{t-5} - 0.047133p_{t-6} - 0.015114p_{t-7} \\
& + 0.000332d_{t-1} + 0.169602d_{t-2} - 0.158671d_{t-3} - 0.005746d_{t-4} \\
& - 0.049974d_{t-5} - 0.011406d_{t-6} + 0.052069d_{t-7} + e_{pt},
\end{aligned}$$

Adjusted $R^2 = 0.990667, \quad$ SE of equation $= 0.039186, \quad$ AIC $= -3.632965,$

$$\begin{aligned} d_t = &-0.034355p_{t-1} + 0.039080p_{t-2} + 0.002444p_{t-3} + 0.004711p_{t-4} \\ &- 0.001891p_{t-5} + 0.011086p_{t-6} - 0.014586p_{t-7} \\ &+ 1.218030d_{t-1} - 0.128928d_{t-2} + 0.007982d_{t-3} - 0.022163d_{t-4} \\ &- 0.120889d_{t-5} + 0.125577d_{t-6} - 0.096529d_{t-7} + e_{dt}, \end{aligned}$$

Adjusted $R^2 = 0.993749, \quad$ SE of equation $= 0.015021, \quad$ AIC $= -5.550691.$

The lag order is selected by the AIC of the overall model which is found to be -9.191136. Based on the Ljung-Box Q-statistics for lags up to 20, we find no evidence that the residuals are serially correlated at the 1-percent level. We further perform the Bai-Perron break test on the squared residuals e_{pt}^2 but detect no structural break. For e_{dt}^2, the estimated break dates are 1905:4 and 1952:3.

5.11 References

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18(1), 1-22.

Blume, L., Easley, D., & O'hara, M. (1994). Market statistics and technical analysis: The role of volume. Journal of Finance, 49(1), 153-181.

Boswijk, H. P., Hommes, C. H., & Manzan, S. (2007). Behavioral heterogeneity in stock prices. Journal of Economic Dynamics and Control, 31(6), 1938-1970.

Brock, W. A., & Hommes, C. H. (1997). A rational route to randomness. Econometrica, 65(5), 1059-1095.

Brock, W. A., & Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic dynamics and Control, 22(8), 1235-1274.

Brock, W. A., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

Brown, D. P., & Jennings, R. H. (1989). On technical analysis. Review of Financial Studies, 2(4), 527-551.

Chen, X., & Ludvigson, S. C. (2009). Land of addicts? An empirical investigation of habit?based asset pricing models. Journal of Applied Econometrics, 24(7), 1057-1093.

Evans, G. W., & Honkapohja, S. (2001). Learning and expectations in macroeconomics. Princeton University Press.

Grundy, B. D., & McNichols, M. (1989). Trade and the revelation of information through prices and direct disclosure. Review of Financial Studies, 2(4), 495-526.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis, 48(5), 1433-1461.

Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica, 50(5), 1269-1286.

Hommes, C. (2013). Behavioral rationality and heterogeneous expectations in complex economic systems. Cambridge University Press.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. Econometrica, 46(6), 1429-1445.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60(7), 1772-1791.

Treynor, J. L., & Ferguson, R. (1985). In defense of technical analysis. Journal of Finance, 40(3), 757-773.

Uhlig, Harald. (2001). A toolkit for analysing non-linear dynamic stochastic models easily. Computational methods for the study of dynamic economies. Oxford University Press, 30-61.

Wang, J. (1993). A model of intertemporal asset prices under asymmetric information. The Review of Economic Studies, 60(2), 249-282.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126.

Zhou, G., & Zhu, Y. (2013). An equilibrium model of moving-average predictability and time-series momentum. Unpublished working paper, Washington University in St. Louis.

Chapter 6

Conclusion

The aim of this thesis is to further tie technical analysis to modern finance theory in an attempt to tighten this gap in the literature. The intended outcome is to promote building a stronger theoretical basis for the use of technical analysis as an investment tool. We have developed four chapters, each as a standalone research paper, to study two portfolio choice problems and two asset pricing problems in which investors make strategic use of information from technical analysis, specifically the moving averages. Our model approach provides several new insights to the field. We summarize the main findings and their implications as follows.

In Chapter 2, we build a model to study the effects of the uncertain predictive power of moving averages on portfolio choice. We find that investors accounting for such uncertainty allocate substantially less to stocks and are more conservative in market timing for longer horizons. Furthermore, the utility loss of ignoring this uncertainty becomes sizable as investment horizon increases. These findings help justify why long-horizon investors seem to ignore much information from technical analysis, while short-horizon investors, who tend to be more speculative, react more strongly even though they know that such information need not be reliable.

In Chapter 3, we present another portfolio choice model to theoretically illustrate that moving averages can be useful for investment when stock returns are correlated. Calibrating our model with S&P 500 price index and dividend yield data, we find that MA-based market timing can substantially improve annualized expected holding period returns. The model also implies that shorter-horizon investors optimally time the market more aggressively and their portfolio profitability is more robust to parameter estimation errors in their return prediction models, results that are consistent with that of Chapter 2.

In Chapter 4, we formulate an asset pricing model and propose some plausible equilibria in which future prices can be predicted by moving averages. This model provides a theoretical basis for some recent empirical findings that moving averages have predictive power. Interestingly, we find that even if prices are determined only by rational fundamental investors who forecast future prices based on information from dividends, due to serial correlations in dividends, technical investors would still find that moving averages have genuine predictive power. We also show that it is possible to have an equilibrium in which prices are determined only by technical investors who forecast future prices based on information from moving averages. However, for this equilibrium to be stable, a relatively short lookback period has to be used to compute the moving averages, implying that the market would have to be relatively informationally efficient in the sense that remote past prices do not affect future prices.

In Chapter 5, we formulate a similar asset pricing model to that of Chapter 4 with special attention to developing estimation and testing strategies to examine whether the proposed equilibria are empirically supported. Using S&P 500 index and dividend data for the period January 1871 to December 2015, we empirically reject the possibility that investors' trend following behaviour, including the use of moving averages, is the driver of the stock market in the long run. Instead, our results support the notion that stock prices reflect fundamental values, despite the widespread use of technical analysis. This finding is consistent with the theoretical result from Chapter 4 that even "rational" prices can imply the predictive power of moving averages.

There are several limitations of our study. Most notably, we have restricted our attention to primarily the moving averages. While moving averages are undoubtedly a representative class of tools in technical analysis, more elaborate trading rules are not considered in this thesis. We are essentially hoping that our moving-average based models give us sufficient insights and confidence to generalize our claims to other technical indicators, which are also trend following in nature.

We are also aware that some model assumptions are not innocuous as they appear. In Chapter 1, we assume that the investor ignores hedging demands from the dynamic learning of the prediction model parameters. Such hedging demands can have important effects on portfolio choice but are neglected due to mathematical complexity. In Chapter 2, we assume that the drift of the stock returns follows the Ornstein-Uhlenbeck (OU) process, and in the calibration exercise we further assume that it is a linear function of log dividend yields (state variable). We acknowledge that the choice of this state variable is arbitrary, for example, alternatives at least include term spreads and payout ratios. This choice of state variable follows from most previous related studies but we do not investigate how the estimated model parameters would change if alternative state variables were used. Besides, whether the dynamics of the log dividend yield can be well described by the OU process is not empirically justified. In an unreported exercise, we find statistical evidence that the log of monthly S&P 500 dividend yields are actually second-order autocorrelated, violating the first-order autocorrelation assumption commonly imposed in previous studies (the OU process implies first-order autocorrelation). For mathematical tractability, we nonetheless ignore such model misspecification and fit data to the OU process because the continuous-time equivalence of the second-order autoregressive model is difficult to work with.

In Chapter 3 and 4, we assume that dividends are stationary and thus the equilibrium prices are also stationary. This stationarity property clear does not hold for most stock market data. We instead detrend both the logs of price and dividend data and interpret the detrended data as observations from an otherwise "real-world equivalent" stationary economy. While similar detrending approaches are common in the macroeconomic literature, the extend to which our testing results are affected by such detrending remains unclear. However, our detrending approach does preserve the observed correlation between the logs of price and dividend data, ensuring no artificial correlation is created.

We also acknowledge that the statistical inference in Chapter 5 is essentially based on the estimation of the risk aversion parameter and it is estimated under the assumption of representative investor with infinite horizon. This assumption is unrealistic but is generally considered a useful simplifying assumption in asset pricing. One can imagine there are infinite generations of investors who share the same risk aversion. Having a sole risk aversion parameter across all generations may still sound very strong, we can further imagine we are seeking an average risk aversion parameter such that it well represents all generations in our sample period. Indeed, the literature on time-varying risk aversion suggests that risk aversion is mean-reverting.

In light of the results of the thesis, a notable feature of the models introduced is the explicit modelling of the interdependency between the stock prices and the investor's forecasts of future prices. We have illustrated this feature in three different contexts. In Chapter 2, while the actual stock price process is exogenous, the individual (non-representative) investor uses his forecast model to speculate how the price may behave in the longer run. Such subjective speculative beliefs have no actual effect to the economy but is critical for the investor's portfolio choice decision. In Chapter 3, since the investor is assumed to have the log-utility preference, he behaves "myopically" and only the next immediate instant of price change is relevant—with no further feedback considered. In Chapters 4 and 5, we demonstrate how the actual process of stock prices and the investor's forecasts can form a tight feedback mechanism.

This thesis opens some possible directions for future work. The following suggestions are illustrative rather than exhaustive. For portfolio choice problems, models that allow for the role of dynamic learning of prediction model parameters is called for. The use of technical analysis in stock markets with time-varying volatility also remains uninvestigated. For asset pricing problems, an equilibrium model allowing for investors' heterogeneous forecasting rules is called for. Estimating the relative proportion of fundamental and technical investors and empirically test such a model are also worth attempting. All of these are important and challenging topics for future research.