

Small levels of predictability and large economic gains

Yexiao Xu*

School of Management, The University of Texas at Dallas, P.O. Box 830688, Richardson, TX 75083, USA

Accepted 7 April 2003

Abstract

Small-return predictability in the stock market has been widely documented in empirical studies, yet little has been written on its economic importance. This paper examines the issue through profitability on a trading strategy that utilizes small levels of predictability and analyzes the statistical distribution for returns achieved under such a strategy. Our results suggest that small-return predictability is economically significant in the sense that such trading strategies not only yield high returns but also are less risky under a fat tail distribution assumption. Quantitatively, we demonstrate that such a strategy could have doubled the market return for the period 1952–1998. We investigate reliability of our results through simulation and bootstrapping.

© 2003 Elsevier B.V. All rights reserved.

JEL classification: C16; G10; G14

Keywords: Conditional expectation; Economic significance; Predictability; Student-*t* distribution; Trading strategy

1. Introduction

Predictability in stock returns and understanding returns behavior are important issues to both academics and Wall Street professionals. Return behavior can also have implications for the efficient market hypothesis, and it can influence asset allocations and capital decisions.

Previous research has found that there are small but statistically significant positive serial correlations in short-term returns series (see, e.g., [Conrad and Kaul, 1988, 1989](#); [Fama, 1970, 1991](#)) and negative serial correlation in long-term return series (see, e.g.,

* Tel.: +1-972-883-6703; fax: +1-972-883-6522.

E-mail address: yexiaoxu@utdallas.edu (Y. Xu).

Fama and French, 1988a; Lo and MacKinlay, 1990a). Other empirical research also suggests that, other fundamental variables, such as interest rates, dividend yields, price–earnings ratio and term structure variables (see Campbell, 1987; Campbell and Shiller, 1988; Fama and French, 1988b; Cutler et al., 1991; Keim and Stambaugh, 1986) are good predictors for stock returns.

Two important issues have emerged, the economic significance of small-returns predictability and its implications for the efficient market hypothesis. We can measure economic importance of small-returns predictability from an *ex ante* perspective. For example, we can interpret the Campbell (1991) study as an investigation of the *ex ante* economic effect of returns predictability. Campbell discusses how predictability can change our understanding about the persistence in expected returns. Kandel and Stambaugh (1996) take an indirect approach, asking whether, given a knowledge of predictability, rational investors' portfolio decisions can be altered. Kandel and Stambaugh show that even statistically weak predictability can have a substantial impact on risk-averse Bayesian investors' portfolio decisions. Barberis (2000) investigates the issue by asking how, in the presence of predictability, estimation risk and investment horizon affect investors' asset allocation. Cochrane (1999) offers a unique survey into the issue.

In this paper, we investigate the economic importance of the return predictability using a direct approach—measuring trading profits. In particular, we test a dynamic trading strategy that is based on updated estimates taken throughout the sample period. Mathematically, we show that such a trading strategy is useful in measuring economic significance of predictability, in the sense that it generates high returns with low risk even under a fat tail distribution assumption such as Student-*t*. Quantitatively, our empirical results suggest that the observed small return predictability not only can help to double the benchmark returns for the period 1952–1998, but also generates a high Sharpe ratio by applying the trading strategy. Since the trading strategy only requires rebalancing the investment portfolio monthly, it is also robust even with transactions costs. We demonstrate the reliability of our results through simulating a return process with an R^2 of 2.5% and bootstrapping on actual returns. There might be institutional details that prevent the realization of economic profits of the active trading strategy in practice. However, this is not very relevant since the goal of this paper is to provide a simple framework to assess the economic importance of predictability. Therefore, by this measure, small levels of predictability are also economically very important.

If predictability is indeed both statistically significant and economically important, we must understand the mechanism that could produce such predictability before we draw any conclusions about the efficient market hypothesis. Return predictability can be attributed to several factors. The positive correlations observed in high frequency returns may come from market microstructure effects, such as the non-synchronous trading and the bid-ask bound, as suggested by Fisher (1966), Boudoukh et al. (1994) Kaul and Nimalendran (1990) and Lo and Mackinlay (1990b). Long-term predictability could be explained by time-varying risk premiums, as shown by Fama (1991), Hirshleifer (1975), LeRoy (1973), Lucas (1978) and Poterba and Summers (1988). For the “intermediate” periods, non-synchronized earning announcement phenomenon could be an important factor.

Different corporations usually end their fiscal quarters in different months, so their earnings announcements tend to occur at different dates. Most empirical results suggest

that earning announcements contain both unexpected firm-specific news and news about the economy and other common factors that can influence all firms. In a world where returns are characterized by a multifactor model, we can anticipate that, if financial markets are efficient, both the price of the relevant stock and also prices of other stocks should adjust immediately to earning announcements. When the common factors indicate a “strong economy,” not only could current earnings of a firm be high, which causes high stock prices, but such good news could also be confirmed by the subsequent earning announcements of other firms. Stock prices will continue to rise as the strength of the economy becomes apparent through the series of earning announcements. That is, a positive return follows yet another positive return. The implied “possibility” is important. Without exogenous imperfections, such as a “noisy supply,” investors can filter out firm-specific news in earning announcements and be able to know the exact state of the economy. However, in reality, there are liquidity traders who prevent a perfect revelation of the true state of the economy.

Also related to this paper is the work by Elliott and Timmermann (2002), Granger and Pesaran (2000) and Pesaran and Timmermann (1995), where they argued that in evaluation of stock market forecasts profits generated from using the forecasts are more appropriate than the conventional statistical measures of forecast accuracy. In particular, Granger and Pesaran (2000) have studied the issue of economic versus statistical measures of forecast accuracy using a two-state Markov process. In contrast, this paper studies a similar problem under a general distribution assumption. From a practical perspective, we also investigate the issue of to what extent we can translate statistical predictability into economic value. The paper is organized as follows. In Section 2, we set the theoretical foundation for studying return predictability in our framework. We then discuss the major proposition of the paper, which demonstrates the feasibility of our trading strategy. In Section 3, we examine the issue of model selection in the context of trading profits. Section 4 investigates the economic significance of predictability. Section 5 concludes.

2. Can small levels of predictability be economically significant?

Researchers have applied different statistical approaches, including autocorrelation, variance ratio and VAR-based, to show returns predictability. Kaul (1996) provides a detailed discussion of econometric issues in testing predictability. The consensus is that to some degree, stock returns are predictable. In fact, it is common to observe moderate R^2 s for major stock indexes using autoregression models. Here, the word “moderate” represents 1% or 2% in R^2 s, which would mean little to an econometrician. The linear models that we often use today cannot perfectly interpret the nonlinear world. Hence, such a change in magnitude of R^2 s may partly be attributed to approximation or sampling errors.

However, in reality, predictability still is important to our profession. Predictions can be used to challenge our beliefs about market efficiency and to influence investors’ decisions. Campbell (1991) shows that small but persistent variations in expected returns can have a dramatic impact on a security’s price. Kandel and Stambaugh (1996) argue from a Bayesian perspective that small levels of predictability could significantly alter a Bayesian

investor's portfolio allocation. Using more direct approaches, researchers have also explored various contrarian strategies. Much attention has been focused on a very short return horizon, such as daily returns (see, e.g., Brock et al., 1992), as well as on a long return horizon, such as several years. Due to substantial transactions costs, it is difficult to explore any profitable strategies that use very short-run predictability. At the same time, there are not enough data to claim the reliability of strategies based on very long-run predictability even without concerning the statistical artifacts of long-horizon returns (see, e.g., Kirby, 1997). To shed light on small-returns predictability, we focus solely on monthly return data.

We intend to use profitability as one of the direct measures of economic significance of predictability. Therefore, we ask if investors can improve their trading profits when they can gauge returns predictability. Thus, we propose the trading strategy as follows:

Trading Strategy $\mathcal{T} = \{\text{Invest in a risky asset today only if the predicted future asset return is positive.}\}^1$

It is well documented that financial data can be best modeled using “fat tail” distributions. Among those distributions, Student- t distribution is widely used. To implement the strategy, suppose return y_t is distributed as Student- t with v degree of freedom. Its unconditional mean and variance are μ and $(v/(v-2))\sigma_1^2$, respectively. Based on information set $\mathcal{F}_{t-1}$ available at time $t-1$, we define the predicted future return as $\hat{y}_t = E(y_t | \mathcal{F}_{t-1})$. When $\mathcal{F}$ has a joint Student- t distribution with the same degree of freedom, the predicted return $\hat{y}_t$ will also be distributed as a Student- t (see Zellner, 1971). For ease of exposition, we concentrate on unbiased predictors in the sense that $E(\hat{y}_{t+1}) = \mu$. In other words, the joint distribution for $y = [y_t, \hat{y}_t]'$ can then be expressed as

$$f(y) = \frac{\Gamma(\frac{v}{2} + 1)}{\Gamma(\frac{v}{2}) \sqrt{v^2 \pi^2 |\Sigma|}} \left[1 + \frac{1}{v} (y - \mu)' \Sigma^{-1} (y - \mu) \right]^{-\frac{v+2}{2}}, \quad (1)$$

where,

$$E(y) = \mu = \begin{bmatrix} \mu \\ \mu \end{bmatrix}, \text{Var}(y) = \frac{v}{v-2} \Sigma = \frac{v}{v-2} \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}, \quad (2)$$

and ρ is the correlation coefficient between actual returns and the predicted returns. To assess the value of knowing predicted $\hat{y}_t$ beforehand, we establish the following proposition:

Proposition 1. When the condition, $\beta = \frac{\text{Cov}(y_t, \hat{y}_t)}{\text{Var}(\hat{y}_t)} \geq \frac{v-1}{v} \frac{\lambda^2}{1+\lambda^2}$, is satisfied, the expected return under trading strategy $\mathcal{T}$ exceeds the unconditional expected return μ , i.e.,

$$E[y_t | \mathcal{T}] = \mu \Psi_v(\lambda) + \mu \frac{\beta}{\lambda} \frac{v + \lambda^2}{v - 1} \psi_v(\lambda) > \mu, \quad (3)$$

¹ When there exists a risk-free asset, excess return can be used instead. In the presence of transactions costs, it can simply be subtracted from the expected return under the trading strategy. This is a conservative approach to deal with transactions costs since trading will not occur at all points of time.

where $\psi_v(\lambda) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}} [1 + \frac{\lambda^2}{v}]^{-\frac{v+1}{2}}$ and $\Psi_v(\lambda) = \int_{-\infty}^{\lambda} \psi_v(w)dw$ are the Student- t density and cumulative t density, respectively, and $\sqrt{\frac{v-2}{v}}\lambda = \frac{\mu}{\sqrt{\frac{v}{v-2}\sigma_2}}$ is the out-of-sample Sharpe ratio. The corresponding volatility is also strictly less than that of the unconditional return volatility with

$$\begin{aligned} \text{Var}[y_t | \mathcal{T}] &= \frac{v\sigma_1^2}{v-2} + \sigma_2^2 \left[\left(\lambda^2 + v \frac{\sigma_1^2/\sigma_2^2}{v-2} \right) [\Psi_v(\lambda) - 1] + \frac{v\beta^2}{v-2} [\Psi_{v-2}(\lambda) - \Psi_v(\lambda)] \right. \\ &\quad \left. + (2\beta - \beta^2) \frac{v\lambda + \lambda^3}{v-1} \psi_v(\lambda) \right] \\ &< \text{Var}[y_t]. \end{aligned} \quad (4)$$

Proof. See Appendix A. □

Proposition 1 gives the theoretical foundation for evaluating the economic significance of return predictability. Since a buy-and-hold trading strategy delivers an unconditional expected return, Eq. (3) implies that the proposed trading strategy $\mathcal{T}$ is superior as long as the actual returns can be predicted to some degree. Moreover, $E[y_t | \mathcal{T}]$ increases monotonically with β , which is in turn proportional to ρ since $\beta = \rho(\sigma_1/\sigma_2)$. The difference, therefore, measures the economic significance of predictability. Similarly, Eq. (4) shows that returns realized under such a trading strategy should also be relatively less risky than that under a buy-and-hold trading strategy.

The proposition also nests the special case of normally distributed returns when the degrees of freedom v approach to infinity. In fact, this case can be derived more directly using the properties of the conditional mean and variance of a truncated normal distribution (see, for example, Greene, 2000; Maddala, 1983). The intuition behind Proposition 1 is straightforward. Due to the positive autocorrelation (or predictability), the joint distribution ellipse between the predicted return and actual return will rotate toward the first quadrant. Meanwhile, we can eliminate a substantial probability mass for predicting a negative return when the true return is indeed negative since return variability is larger than the mean. This elimination leaves the conditional distribution positively skewed. Therefore, the conditional mean conditioned on predicting a positive return will increase.

The condition in Proposition 1 is very weak. In general, we can write $y_t = \hat{y}_t + \epsilon_t$. If $\hat{y}_t$ is an in-sample predictor, it is orthogonal to ϵ_t . In this case, the proposition condition naturally holds, since $\beta=1$. However, because we use an out-of-sample predictor, $\hat{y}_t$, it is not necessarily orthogonal to ϵ_t . Despite that, the condition easily holds as shown in Section 3.

We can rewrite the condition in terms of the predictive power of $\hat{y}_t$ for y_t (i.e., the coefficient of determination) in the following form,

$$R^2 \geq \left[\frac{\frac{v-1}{v-2} \eta_y \eta_{\hat{y}}}{1 + \frac{v}{v-2} \eta_{\hat{y}}^2} \right]^2, \quad (5)$$

where $\eta_y = \frac{\mu}{\sqrt{\frac{v}{v-2}\sigma_1}}$ is the Sharpe ratio for the original returns, and $\eta_{\hat{y}} = \frac{\mu}{\sqrt{\frac{v}{v-2}\sigma_2}}$ is the Sharpe ratio for predicted returns (or the out of sample Sharpe ratio). Clearly, condition (5) suggests that we do not need to have large magnitude of predictability in returns when v is large. In

other words, kurtosis in return will have some impact in achieving the economic significance of predictability.

To analyze the relative performance of the proposed trading strategy $\mathcal{T}$, we define the efficiency gain as $q(\cdot) = E[y_t | \mathcal{T}] / E[y_t]$. This gain can be computed from Eq. (3), i.e., $q(\beta, \lambda, \nu) = \Psi_\nu(\lambda) + \frac{\beta}{\lambda} \frac{\nu + \lambda^2}{\nu - 1} \psi_\nu(\lambda)$, which is a function of β , λ and ν . For comparison, in Panel

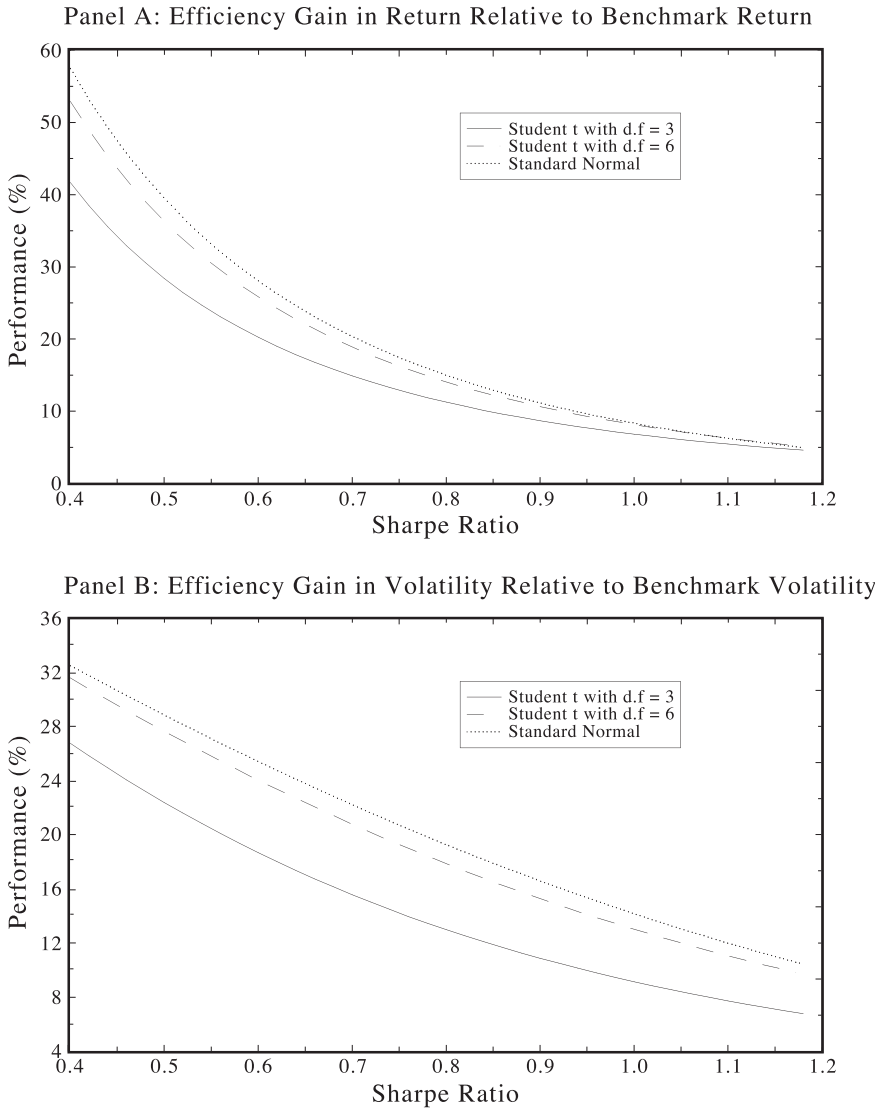


Fig. 1. The relative performance of expected return and volatility. This graph shows the performance gain in the expected return and decrease in the volatility of returns using the active trading strategy. “Sharpe ratio” stands for the out of sample Sharpe ratio η . I have used $\beta=1$ and $\rho=R^2$ to plot the graph.

A of Fig. 1, we plot the relation between efficiency gain and the out-of-sample Sharpe ratio ($\eta_{\hat{y}}$) for different degree of freedoms when $\beta=1$. The figure shows that the efficiency gain from the trading strategy $\mathcal{T}$ is inversely related to λ . This is also implied by Lemma 1 in Appendix A. Since the strategy should work well when the out-of-sample predicted returns capture most of the variations in the actual returns, i.e. small out-of-sample Sharpe ratio $\eta_{\hat{y}} (= \sqrt{\frac{v-2}{v}}\lambda)$, this result is intuitive. For example, under normality assumption, when $\eta_{\hat{y}}$ is 0.6, the expected return from the trading strategy increases by 29%. At a Sharpe ratio of 0.8, the expected return is 16% higher than that of a buy-and-hold strategy. The efficiency gains are much smaller when returns have a fat tail distribution. In the extreme case where kurtosis does not exist, i.e., $v=3$, the efficiency gain reduces to 21% for $\eta_{\hat{y}}=0.6$. Therefore, we can consider the solid line in Panel A to be the lower bound for the efficiency gain. For a more realistic scenario with $v=6$, which corresponds to a kurtosis of 6, the efficiency gains are close to that of normal distribution.

Similarly, we can also investigate the relative reduction in return volatility. The high volatility in the predicted returns results in a out of sample low Sharpe ratio. This justifies an inverse relation between volatility and λ as shown in Panel B of Fig. 1. Unlike the returns, the volatility lines converge much slower for different degree of freedoms when the λ s are large over the range shown. Using the graph, we can also compute the expected Sharpe ratio for the realized return under $\mathcal{T}$. For example, when $\eta_{\hat{y}}=0.6$, the Sharpe ratio for the trading strategy is 1.82 and 1.74 times that of a buy-and-hold strategy for normally distributed returns and t distributed returns with degree of freedom of 6, respectively.

3. Return structure and model selection

Proposition 1 states that, even if returns are generated from a fat tail distribution, such as the Student- t , the trading strategy $\mathcal{T}$ can be used to demonstrate the economic significance. Clearly, the magnitude of gains from the strategy relative to a buy-and-hold strategy depends on the level of predictability. Therefore, we must put more structure on returns. Since some evidence of predictability stems from return autocorrelations, we assume without loss of the generality that the true data generating process (DGP) for return y_t follows an AR(2) process in the following form:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t \quad (6)$$

where e_t is distributed as i.i.d. Student- t with mean zero, variance σ^2 and degree of freedom v . Similar to a normal distribution, Zellner (1971) has shown that a linear combination of t distributed random variables with v degree of freedom is still distributed as a t with the same degree of freedom. Therefore, any predicted value of $\hat{y}_t$ based on an AR(p) model will be unbiased with a Student- t distribution of v degree of freedom.

In reality, we do not know either the true return generating process (an AR(2) in this case) or the true model parameters. We must first select an appropriate model, and then estimate model parameters from sample data. To study the robustness of these results and the quantitative efficiency gain, we construct three typical cases of model selection and sample size.

3.1. Case I: large samples with an AR(1) model

Since the true DGP given in Eq. (6) is unknown, we postulate the following AR(1) model:

$$y_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + u_{t+1}. \quad (7)$$

Assuming a large sample, the standard asymptotic results hold due to the i.i.d. assumption. The least square estimators $\hat{\phi}_0$ and $\hat{\phi}_1$ converge to $(\phi_0/(1-\phi_2))$ and $(\phi_1/(1-\phi_2))$, respectively. Although these parameter estimates are upward-biased by a factor of $(1/(1-\phi_2))$, the predicted return $\hat{y}_t (= \hat{\phi}_0 + \hat{\phi}_1 y_{t-1})$ is unbiased because the estimated residuals have zero means. Moreover, if the variances of estimators are close to zero in a large sample, we have $\text{Var}(\hat{y}_{T+1}) \approx \text{Cov}(\hat{y}_{T+1}, y_{T+1})$. In other words, when we can estimate the model parameters very accurately, β is close to one, which means Proposition 1 applies.

This analysis demonstrates that in a large sample, we can still achieve economic profits with the trading strategy even though we assume an incorrect model. Asymptotically, since $\hat{y}_t$ has a t distribution with $(\nu/(\nu-2))\sigma_2^2 = \text{Var}[\hat{y}_t] = \theta^2 \gamma_0$ where $\theta = \phi_1/(1-\phi_2)$ and $\gamma_0 = \text{Var}(y_t)$, the out-of-sample Sharpe ratio can be expressed as $\eta_{\hat{y},I} = \frac{\mu}{\sigma} \sqrt{(1-\phi_2^2)(\theta^{-2}-1)}$.

In order to obtain a numerical assessment of the economic significance of return predictability, we assume $\mu=0.663\%$, $\sigma=4.2\%$, $\phi_1=0.15$, $\phi_2=0.03$ and $\nu=6$ according to the historical log real monthly NYSE/AMEX/NASDAQ composite index returns for the postwar period from 1952 to 1998. This corresponds to a low R^2 of 2.54% and a kurtosis of 6 for the return variable. With these estimates, we compute $\eta_{\hat{y},I}$ to be 1.008. This yields to a 7.97% increase in the expected return and a 12.90% decrease in the volatility, or a 23.96% increase in the Sharpe ratio.

3.2. Case II: large samples with an AR(2) model

If we can choose a correct model that is the same as the true DGP, we can estimate the following AR(2) model:

$$y_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} + u_{t+1}. \quad (8)$$

Since the model is correctly specified, the OLS estimators $\hat{\phi}_0$, $\hat{\phi}_1$ and $\hat{\phi}_2$ are asymptotically unbiased and consistent. Thus, the predictor $\hat{y}_t (= \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2})$ is also asymptotically unbiased. Moreover, since the least square parameter estimates are very close to their true values in a large sample, we obtain $\beta \approx 1$. In this case, Proposition 1 holds for the t distributed predicted return $\hat{y}_t$. In other words, we have the same conclusion as if we know the true model parameters when a large sample is available. Because $(\nu/(\nu-2))\sigma_2^2 = \text{Var}[\hat{y}_t] = \gamma_0 - \sigma^2$, by applying Eq. (B.1), we can write the Sharpe ratio as $\eta_{\hat{y},II} = \eta_{\hat{y},I} [1 + (\phi_2/\phi_1)^2 (1-\phi_2)^2 (1-\theta^2)]^{-1/2}$. Therefore, the out-of-sample Sharpe ratio is smaller than that in the first case. In other words, in a large sample, it pays to use a more complex model. Using the same estimates as in Case I, we find η_{II} to be 0.989. Again, from Fig. 1, the efficiency gain in the expected return is 8.38%, while the reduction in the volatility is 13.29%. This suggests an increase in the Sharpe ratio of 25%. These numbers are slightly larger than those of the previous case.

3.3. Case III: a moderate sample size

In practice, we only expect to have a moderate sample size, which means all models are estimated with errors. Although β will be less than, but close to, one, Proposition 1 condition will still hold. This is because the volatility of stock returns is usually much larger than the mean of stock returns. Even when the volatility of predicted returns is comparable to the average return, the bound on the condition of the proposition is close to a half. Therefore, the trading strategy still outperforms a buy-and-hold strategy. Since researchers have the freedom to choose a model if there is no prior knowledge about the true DGP, we want to assess the relative efficiency gain associated with each model for moderate sample size. We ask which is the best for a moderate sample size, a simple AR(1) model or an AR(p) model.

Although we should always use a model that is consistent with the true DGP for a large sample, such a model may not be appropriate if the sample size is moderate and if the parameters are very small. This is exactly the case since in practice both the return predictability and the magnitude in ϕ_1 and ϕ_2 are very small. It is important to compare the efficiency gains under an AR(1) model and under an AR(2) model when the true DGP follows an AR(2) process.

3.3.1. Model 1: AR(1)

We model monthly returns as an AR(1) process of Eq. (7). From Appendix B, we show that:

$$\text{Var}[\hat{\phi}_0] = \frac{1}{T}(\gamma_0 + \mu^2)(1 - \theta^2), \quad (9)$$

$$\text{Var}[\hat{\phi}_1] = \frac{1}{T}(1 - \theta^2), \quad (10)$$

$$\text{Cov}[\hat{\phi}_0, \hat{\phi}_1] = -\frac{1}{T}\mu(1 - \theta^2), \quad (11)$$

where $\theta = \phi_1/(1 - \phi_2)$. Because of the weak autocorrelation, we assume that the estimators $\hat{\phi}_0$ and $\hat{\phi}_1$ based on the first T observations are independent of y_T and ϵ_{T+1} . By applying Eq. (9) to Eq. (11), we obtain:

$$\begin{aligned} \text{Var}[\hat{y}_{T+1}] &= \text{Var}[\hat{\phi}_0] + (\mu^2 + \gamma_0)\text{Var}[\hat{\phi}_1] + 2\mu\text{Cov}[\hat{\phi}_0, \hat{\phi}_1] + \theta^2\gamma_0 \\ &= \frac{2\sigma^2}{T(1 - \phi_2^2)} + \theta^2\gamma_0, \text{ and} \end{aligned} \quad (12)$$

$$\text{Cov}[y_{T+1}, \hat{y}_{T+1}] = \theta^2\gamma_0. \quad (13)$$

There are $T=564$ months in the postwar period from 1952 to 1998. Using the same parameters as in Case I, we estimate the return volatility to be $\gamma_0=1.809 \times 10^{-3}\%$. Substituting these parameter values into Eqs. (12) and (13), we have the volatility for the predicted return as well as the covariance between predicted return and the actual return as:

$$\text{Var}[\hat{y}_{T+1}] = 4.952 \times 10^{-5} \text{ and } \text{Cov}[y_{T+1}, \hat{y}_{T+1}] = 4.326 \times 10^{-5},$$

respectively. β and λ are thus equal to 0.8736 and 0.9421, respectively. Applying Eq. (3) in Proposition 1, we obtain $E[y_{T+1} | \mathcal{T}] = 0.706\%$, which exceeds the returns that would be realized under the buy-and-hold strategy by 6.5% per month. At the same time, the return volatility under the trading strategy, $\text{Var}[y_{T+1} | \mathcal{T}]$, is only 1.559×10^{-3} , which is 13.8% less than γ_0 . Taking into consideration both increases in the expected return and reductions in volatility, we have a Sharpe ratio of 0.1787, which is 13.2% higher than that under the buy-and-hold strategy. It is also informative to compute a cumulative return. Roughly speaking, using our trading strategy, a dollar invested in 1952 would have returned US\$54 in 1998 in contrast to a return of US\$42 under the buy-and-hold strategy.

3.3.2. Model 2: AR(2)

Omitting a relevant explanatory variable will bias estimators as in Case I of the previous subsection. For this reason, one may tend to include more explanatory variables in order to reduce or to eliminate such biases. However, over-fitting can inflate the variances of estimators, which will substantially increase the variance of predicted returns. Therefore, a more interesting question is whether it is beneficial to fit a model exactly. In this section, we show that even if we use the same model as the true DGP, the expected return could be worse than that based on the AR(1) model. When the model parameters are very small, the gains in the explanatory power by estimating an extra parameter might be less than the error it introduces.

To illustrate the point, we construct an AR(2) model as shown in Eq. (8). The model has the same functional form as the true DGP. The corresponding variances and covariances can be expressed as:

$$\text{Var}[\hat{\phi}_0] = \frac{\sigma^2}{T} \left(1 + \frac{2\mu^2}{\gamma_0(1+\theta)} \right), \quad (14)$$

$$\text{Var}[\hat{\phi}_1] = \text{Var}[\hat{\phi}_2] = \frac{\sigma^2}{T} \frac{1}{\gamma_0(1-\theta^2)}, \quad (15)$$

$$\text{Cov}[\hat{\phi}_0, \hat{\phi}_1] = \text{Cov}[\hat{\phi}_0, \hat{\phi}_2] = -\frac{\sigma^2}{T} \frac{\mu}{\gamma_0(1+\theta)}, \quad (16)$$

$$\text{Cov}[\hat{\phi}_1, \hat{\phi}_2] = -\frac{\sigma^2}{T} \frac{\theta}{\gamma_0(1-\theta^2)}, \quad (17)$$

Again, assuming that the estimators $\hat{\phi}_0$, $\hat{\phi}_1$ and $\hat{\phi}_2$ are independent of y_T and ϵ_{T+1} , we derive:

$$\begin{aligned}\text{Var}[\hat{y}_{T+1}] &= \text{Var}[\hat{\phi}_0] + (\mu^2 + \gamma_0)(\text{Var}[\hat{\phi}_1] + \text{Var}[\hat{\phi}_2]) + 2\mu\text{Cov}[\hat{\phi}_0, \hat{\phi}_1 \\ &\quad + \hat{\phi}_2] + 2(\mu^2 + \gamma_1)\text{Cov}[\hat{\phi}_1, \hat{\phi}_2] + (\phi_1^2 + \phi_2^2)\gamma_0 + 2\phi_1\phi_2\gamma_1 \\ &= \frac{3\sigma^2}{T} + (\gamma_0 - \sigma^2),\end{aligned}\quad (18)$$

$$\text{Cov}[y_{T+1}, \hat{y}_{T+1}] = \gamma_0 - \sigma^2. \quad (19)$$

Applying the same parameter value as before, we have

$$\text{Var}[\hat{y}_{T+1}] = 5.0423 \times 10^{-5}, \text{ and } \text{Cov}[y_{T+1}, \hat{y}_{T+1}] = 4.484 \times 10^{-5}.$$

Although the covariance between the predicted value and the actual value has increased by 3.7%, the volatility in the predicted value has gone up by as much as 10% when compared to that of the AR(1) model. For this reason, both β and λ have decreased to 0.8270 and 0.9003, respectively. When we use Eq. (3) to compute the expected return under trading strategy $\mathcal{T}$, we obtain 0.703% per month, which is lower than that from an AR(1) model. Similarly, the return volatility under the trading strategy is 1.543×10^{-3} , or 14.7% smaller than the original return volatility. The Sharpe ratio of 0.1791 is slightly higher.

Our trading strategy $\mathcal{T}$ is not only useful in measuring the economic significance of a very small return predictability ($R^2=2.54\%$), but also demonstrate that a simple model is more efficient from an econometric estimation perspective. Again, the reason that the conditional expected value is smaller than that under an AR(1) model is that when ϕ_2 is very small, the benefit of reducing biases by including an extra regressor is outweighed by the cost of increasing variability in the estimates. Over the sample period, there is a 6.49% (6.03%) increase in the expected return per month under the AR(1) (AR(2)) model or a 28.6% (25.5%) increase in the cumulative returns. In both cases, return volatilities are about 14% less than those under the buy-and-hold strategy.

3.4. The impact of a fat tail distribution and the sample size

In a practical application, the number of observations available for estimation changes when time progresses. It will be studied in Section 4 through simulation. As a related issue, it is also interesting to know when a sample size is large enough to have the benefit of using an AR(2) model. In Panel A of Fig. 2, we show the expected returns from the active trading strategy using both AR(1) and AR(2) models. When the sample size grows

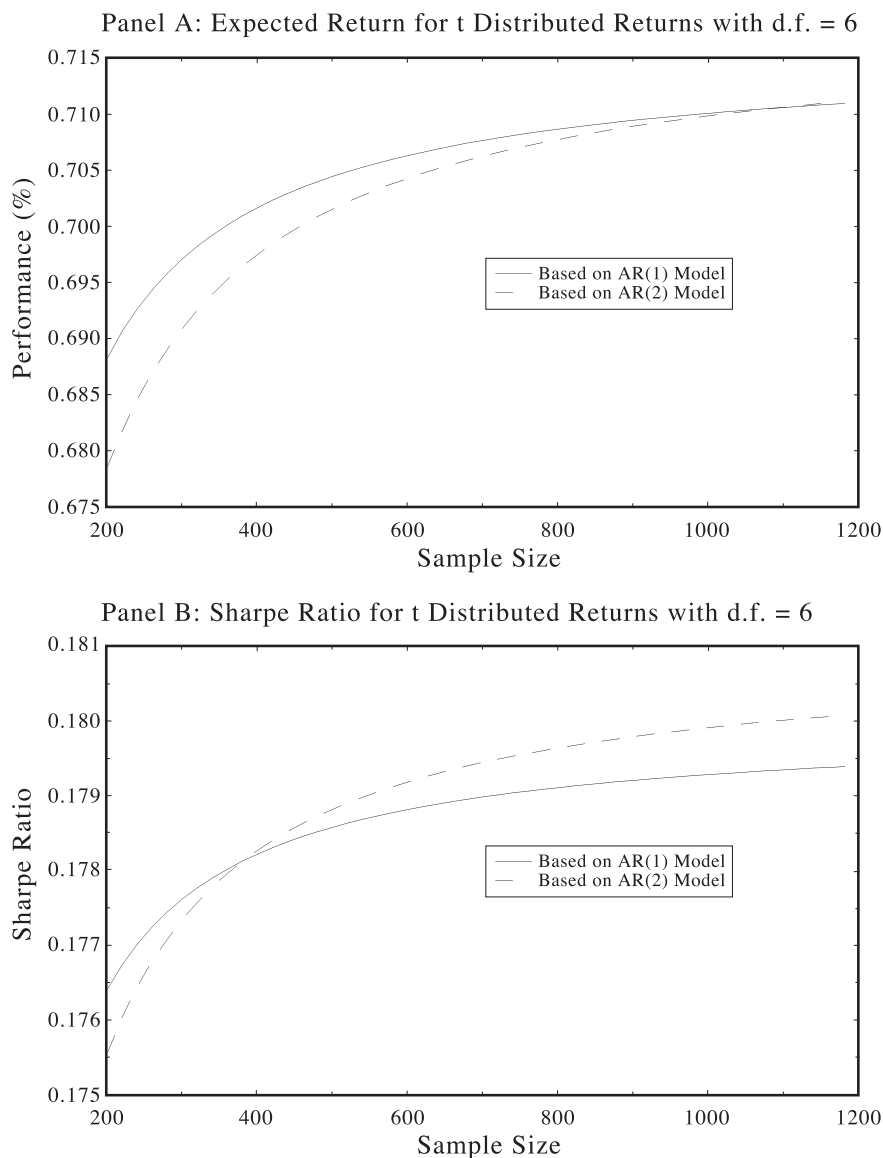


Fig. 2. The relation between performance and sample size. This graph displays the expected returns and Sharpe ratios using both AR(1) model and AR(2) model for different sample sizes. We assume the true data generating process is an AR(2) process with $\mu=0.663\%$, $\sigma=4.2\%$, $\phi_1=0.15$, $\phi_2=0.03$ and t distributed residuals.

beyond 1100 observations, returns under the AR(2) model begin to exceed returns under the AR(1) model. For practical applications using monthly data, we do not have such a large sample size. In other words, we are likely dealing with moderate sample sizes where

an AR(1) model is preferred. The difference between the two models can be substantial. For example, when the sample size is 200, there is about one basis point difference in the expected returns per month.

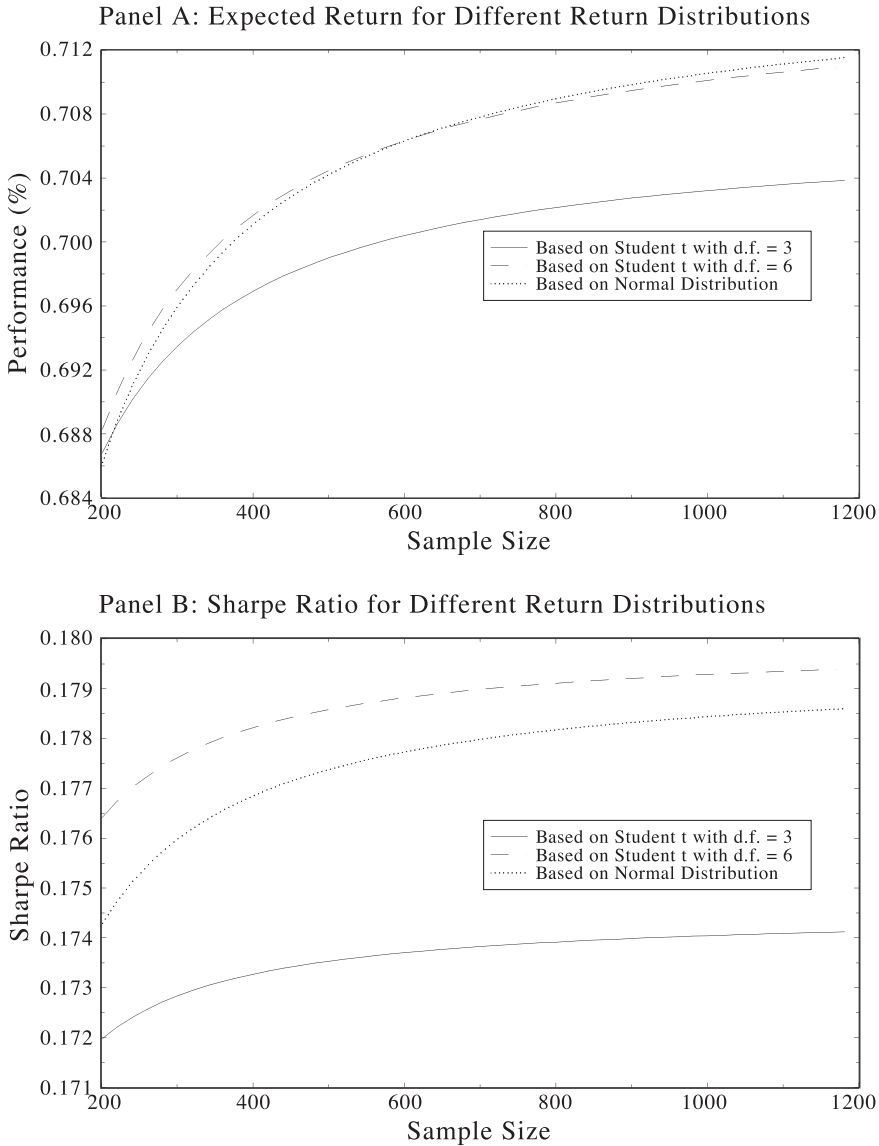


Fig. 3. The relation between performance and sample size for different distributions. This graph displays the expected returns and Sharpe ratios using AR(1) model for different sample sizes. We assume that the true data generating process is an AR(2) process with $\mu=0.663\%$, $\sigma=4.2\%$, $\phi_1=0.15$, $\phi_2=0.03$ and using both Student- t and normal distributions.

However, the total volatility for the return under the active trading strategy is always lower under the AR(2) model than that under the AR(1) model. Therefore, the switching point for selecting a model might be much smaller when both the expected return and the volatility are considered. In Panel B of Fig. 2, we have also plotted the Sharpe ratio. When the sample size is less than 400, the Sharpe ratio obtained from an AR(1) model dominates that from the AR(2) model. The difference in Sharpe ratio is less than 0.0002 for a sample size of 500. In other words, an AR(1) model is still appropriate for most applications.

Return distributions can also be important in affecting the performance as shown in Fig. 1. In the current setup, we compare the normal distribution with the Student- t distribution in Fig. 3. In general, excess kurtosis diminishes gains from implementing the trading strategy for a large sample size. It is also interesting to note that, however, moderate kurtosis can also help enhancing the performance for small to moderate sample size. For example, Panel A of Fig. 3 suggests when excess kurtosis is 3 (i.e., $df=6$) and the sample size is smaller than 400, the performance is better for t distributed returns than for normally distributed returns. A similar pattern holds when performance is measured by Sharpe ratio as shown in Panel B of Fig. 3. Finally, all else equal, the efficiency gain will be large if the original return volatility is high or the original return is low (not shown in the graph).

4. Empirical results

The above theoretical analysis suggests that the proposed trading strategy is useful to measure the economic significance of small levels of predictability in stock returns. Although we are able to focus our discussion so far on the scenario of Student- t distributed stock returns, it is important to consider transactions costs and more general return structures, especially the actual realized returns. In order to accomplish that, we investigate the long-term performance of the proposed trading strategy based on the out-of-sample prediction for the postwar period.

The results will not be very convincing if we focus on the pre-war period, because return predictability is especially strong during that period. The postwar period begins in January 1952, the month that marks the change in the interest rate regime that resulted from the Accord of the Treasury and Federal Reserve in 1951 (see Campbell, 1991). To obtain reliable estimates, we use 10 years of monthly returns (120 observations) before 1952 to start the estimation process. That is, $t=120$ corresponds to December 1951. At any time t , we estimate model parameters based on the first t observations. We then make an investment decision based on the forecasted return, i.e., invest if and only if the forecasted future return is greater than the risk-free rate. Then we advance one period and repeat the process. This procedure implements the proposed trading strategy and, whenever possible, uses all the information.

The historical return sample is just one possible draw from the population distribution. What we observe here may not repeat itself in the future. To ensure the robustness of the study, we first apply the Monte Carlo simulation method to investigate the reliability of using the trading strategy to study economic significance of predictability under different return distributions and transactions costs.

4.1. A Monte Carlo simulation

Applying the same parameterization as in the previous section, we first generate residuals with a normal distribution or a Student- t distribution. Together with the initial condition of $y_0=\mu$, we then create 684 log-return according to the DGP of Eq. (6). We generate the first 120 observations to ensure that we have enough observations to estimate the model parameters at any point of time.

At any time t , we first estimate the parameters of an AR(1) or AR(2) model using available returns up to t . We then forecast a one-period-ahead return based on the estimated model. If the forecasted return is above the risk-free rate, we invest. Otherwise, we hold a risk-free bond that returns 20 basis points each month. This process is repeated 1000 times. For comparison, we first study the case of normally distributed real log returns. The results for our trading strategy using both an AR(1) and AR(2) models are reported in Table 1.

Table 1
Comparison of different trading strategies based on simulated return generated by using normal residuals

	Buy and hold	$R_f=0\%$		$R_f=0.2\%$					
		No transactions costs				With 10 monthly BP		With 20 monthly BP	
		AR(1)	AR(2)	AR(1)	AR(2)	AR(1)	AR(2)	AR(1)	AR(2)
Mean return	0.664	0.685	0.682	0.721	0.714	0.692	0.685	0.662	0.657
(S.D.)	(0.0003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Volatility	4.135	3.818	3.785	3.638	3.602	3.640	3.604	3.643	3.607
(S.D.)	(0.0036)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)
Return difference, t -ratio		0.021, 7.449	0.018, 6.488	0.057, 18.69	0.050, 16.93	0.028, 9.388	0.021, 7.407	−0.001, −0.660	−0.007, −2.62
Average R^2 (S.D.)		2.585 (0.046)	2.944 (0.048)	2.585 (0.046)	2.944 (0.048)				
Cumulative ret.	41.31	46.66	45.95	57.40	55.23	48.44	46.76	40.86	39.56
Sharpe ratio (S.D.)	0.161 (0.0002)	0.180 (0.0009)	0.181 (0.0009)	0.200 (0.001)	0.199 (0.001)	0.191 (0.001)	0.191 (0.001)	0.182 (0.001)	0.183 (0.001)
% of times better than “buy and hold”		59.40	57.80	72.10	69.90	61.10	58.30	49.00	44.60
% of AR(2) better than AR(1)			46.70		40.80		40.50		41.60

This table compares average returns using different trading strategies based on monthly real log stock returns generated from the following process:

$$y_t = 0.0054366 + 0.15y_{t-1} + 0.03y_{t-2} + \epsilon_t$$

where $y_{-1}=y_0=\mu=0.00663$ and ϵ_t is distributed as $N(0, 0.042^2)$. I repeat the simulation 1000 times for each case. Each repetition generates 684 observations. I use the first 120 observations to start the initial estimation of the model parameters. The remaining 564 observations simulate monthly observations from 1952 to 1998. The active trading strategy requires updating model estimates each month and making one-period-ahead forecasts. If the forecasted return is positive, I invest in stock for the next period. Otherwise, I invest in a risk-free bond that returns R_f next period.

For the AR(1) model, with a very small return predictability of R^2 being 2.6% and a risk-free rate of 20 basis points per month, column 5 reveals a striking fact: Return from our simple trading strategy exceeds a buy-and-hold return by about 9% (up from 0.664% to 0.721%) each month, or 39% for the past 47 years. By any measure, this is economically significant. Since we adjust our portfolio once a month at most, transaction costs will not swamp such sizable returns.

Table 1 also shows the corresponding results, with 0.1% and 0.2% monthly transactions costs, respectively.² The return for the active trading strategy is still impressive, a 17% increase from the buy-and-hold return even with a 0.1% transactions costs. However, the return difference between the active strategy and the buy-and-hold strategy disappears when transactions costs increase to 0.2% a month. Since all the return series are highly correlated, we investigate the statistical significance of the results discussed above based on the return difference in each iteration. Table 1 shows that there are about six basis points difference with a t -ratio of 19. Similarly, there are about three basis points difference when applying 0.1% transactions costs.

We could be running risks in implementing the strategy, because it is impossible to guarantee superior performance period by period. In fact, about a quarter of times the buy-and-hold returns are higher. The best way to assess the risk involved is to compare the volatilities of the returns. Clearly, utilizing information on predictability also helps to considerably reduce the return volatility by about 12%, from 4.135% to 3.638%. Moreover, underperformance most often occurs when the predictability is close to zero. Therefore, we should use the Sharpe ratio as an alternative measure. We find that compared to a buy-and-hold strategy, the active strategy provides a Sharpe ratio of 0.200, which reflects more than a 24% increase. Even with 20 bases points of transactions costs, the Sharpe ratio is still 0.182, which represents a 13% increase over that from a passive strategy.

In addition, Table 1 reports results for an AR(2) model under each scenario. The average monthly log real return is 0.721% under the AR(1) model and the AR(2) model produces 0.714%. This result reconfirms our claim that, for a moderate sample size, a simple AR(1) model outperforms an AR(2) model, even though the AR(2) model is the DGP. In fact, an AR(1) model outperforms an AR(2) model 59% of the time. The same pattern prevails even with transactions costs. At the same time, we note that the Sharpe ratio is very close for both models.

Many researchers have found that returns are leptokurtic. For monthly real log-returns, the excess kurtosis is about three. As suggested in the previous theoretical analysis, the Student- t distribution can incorporate such a statistical property very well. By applying the estimation method proposed by McDonald and Xu (1995), we estimate the degree of freedom (df) for the t distribution as 6.4. To generate the corresponding random number from a normal distribution, we use $df=6$, which corresponds to an excess kurtosis of three.

In Panel A of Table 2, using the same procedure to perform simulations, we show results based on t distributed log-returns. These results are very similar to those of Table 1,

² In reality, there will be about 20 bases points charge when holding an index fund over a year. Therefore, the 10 basis points of monthly transaction costs (or 1.2% annual costs) assumed here is very conservative.

Table 2

Comparison of different trading strategies based on simulated return generated by using t residuals

	Buy and hold	$R_f=0\%$		$R_f=0.2\%$					
		No transactions costs				With 10 monthly BP		With 20 monthly BP	
		AR(1)	AR(2)	AR(1)	AR(2)	AR(1)	AR(2)	AR(1)	AR(2)
<i>Panel A: with centralized t residuals</i>									
Mean return	0.663	0.692	0.687	0.720	0.713	0.692	0.686	0.665	0.658
(S.D.)	(0.0003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Volatility	4.137	3.852	3.820	3.683	3.644	3.685	3.646	3.688	3.649
(S.D.)	(0.0055)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)
Return difference, <i>t</i> -ratio		0.029, 10.13	0.024, 8.359	0.057, 18.90	0.050, 16.85	0.029, 10.05	0.023, 7.839	0.017, 0.596	−0.005, −1.63
Average R^2		2.588	2.909	2.588	2.909				
(S.D.)		(0.047)	(0.048)	(0.047)	(0.048)				
Cumulative ret.	41.08	48.47	47.07	56.91	54.78	48.60	46.82	41.48	39.99
Sharpe ratio	0.161	0.180	0.181	0.197	0.197	0.189	0.189	0.181	0.182
(S.D.)	(0.0002)	(0.0009)	(0.0009)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% of times better than “buy and hold”		63.30	61.50	72.40	69.40	62.80	59.80	50.80	48.40
% of AR(2) better than AR(1)			44.90		42.00		42.50		43.00
<i>Panel B: with non-centralized t residuals</i>									
Mean return	0.663	0.707	0.701	0.733	0.728	0.705	0.701	0.678	0.674
(S.D.)	(0.0003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Volatility	4.147	3.851	3.820	3.696	3.660	3.699	3.662	3.702	3.665
(S.D.)	(0.0053)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Return different, <i>t</i> -ratio		0.044, 14.79	0.039, 13.26	0.070, 22.78	0.065, 21.50	0.043, 14.27	0.038, 12.94	0.015, 5.232	0.012, 3.977
Average R^2		2.572	2.912	2.572	2.912				
(S.D.)		(0.046)	(0.048)	(0.046)	(0.048)				
Cumulative ret.	40.96	52.78	51.21	61.28	59.54	52.37	51.08	44.73	43.79
Sharpe ratio	0.160	0.185	0.185	0.200	0.200	0.192	0.193	0.185	0.185
(S.D.)	(0.0002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% of times better than “buy and hold”		66.00	65.00	76.20	73.60	66.50	63.80	55.50	54.20
% of AR(2) better than AR(1)			40.80		43.50		45.00		46.6

This table compares average returns using different trading strategies based on monthly real log stock returns generated from the following process:

$$y_t = 0.0054366 + 0.15y_{t-1} + 0.03y_{t-2} + \epsilon_t$$

where $y_{-1}=y_0=\mu=0.00663$. In Panel A, ϵ_t is distributed as centralized Student- t with 6 degrees of freedom. While in Panel B, ϵ_t is distributed as non-centralized Student- t with 7 degrees of freedom and a non-centralized parameter of -1.0 . The distribution has also been renormalized to yield mean zero and a standard deviation of 0.042. I repeat the simulation 1000 times for each case. Each repetition generates 684 observations. I use the first 120 observations to start the initial estimation of the model parameters. The remaining 564 observations simulate monthly observations from 1952 to 1998. The active trading strategy requires updating model estimates each month and making one-period-ahead forecasts. If the forecasted return is positive, I invest in stock for the next period. Otherwise, I invest in a risk-free bond that returns R_f next period.

but show a slightly better performance in terms of mean return. It seems that with a high probability of extreme values, our trading strategy can prevent large negative events while preserving the positive extremes. Therefore, the presence of excess kurtosis in returns can also improve our results especially for moderate sample size.

To put our results into perspective, we also perform simulations for no transactions costs and a zero risk-free rate. Table 2 reports the results in column 3 for the AR(1) model and in column 4 for the AR(2) model. The results from our theoretical analysis in the previous section (0.706% for the AR(1) model and 0.703% for the AR(2) model) are higher than the simulation results reported here (0.692% for the AR(1) model and 0.687% for the AR(2) model). This is expected since the sample size gradually increases over time in each replication of the simulation. The above analysis indicates that the active strategy does better when the sample size increases. We also see that the volatility is about 7% lower and the Sharpe ratio is 12% higher than those from the buy-and-hold strategy.³

The monthly returns are also negatively skewed. To account for the skewness, we use a noncentralized t distribution with degree of freedom being seven and a noncentralized parameter $\delta = -1$. With such parameters, we obtain a kurtosis of three and a skewness of -0.6 . These results match the properties of the empirical distribution. We can renormalize the random numbers thus generated to have zero mean and the assumed standard deviation. Panel B of Table 2 presents the simulated results. The returns from our trading strategy are much better than those shown in Table 1. The cumulative returns are now 50% higher with no transactions costs, and even 10% higher with 0.2% monthly transactions costs, than the results under a buy-and-hold strategy. This exercise demonstrates that in measuring economic significance of return predictability, skewness in returns can also be very important. Negative skewness generally means that large negative returns are more likely to occur than are large positive returns. If investors care more about downside risk than upside potential, negatively skewed returns might be much riskier than standard deviation would indicate. In this sense, the active trading strategy is very effective in preventing such downward risk.

4.2. Results using actual returns

The return performance from our simulations demonstrates the economic significance of small return autocorrelations. These results are also robust under different distributional assumptions and transactions costs. However, the actual stock returns have a more complex distribution structure than a parametric distribution function can capture. Moreover, return predictability is not due to return autocorrelations alone. Predictability can change over time, in a similar way as volatility does. Therefore, we study the economic significance issue based on real stock returns. Because it is impractical to test our trading strategies in an ongoing stock market, we perform an out-of-sample test on the historical realized return data.

³ The buy-and-hold strategy would have been optimal if returns were generated from a random walk model. This exercise is intended to show the differences in magnitude when returns do not follow a random walk model.

There are other factors such as dividend yield (see Shiller, 1981; Fama and French, 1988b, 1989) and relative interest rates (Campbell, 1987) contributing to return predictability in addition to serial correlations in stock returns. To be comparable to other studies, we use real log returns. We deflate these returns by the inflation rate shown in the 1999 edition of *Ibbotson's Yearbook*. Since inflation rate is persistent, we can use it as additional predictor. We investigate the following models:

$$\text{Model I: } y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \text{inf}_{t-1} + u_t, \quad (20)$$

$$\text{Model II: } y_t = \beta_0 + \beta_1 y_{t-1} + \beta_3 \text{rrf}_{t-1} + \beta_4 \frac{D}{P} + u_t, \quad (21)$$

$$\text{Model III: } y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \text{inf}_{t-1} + \beta_3 \text{rrf}_{t-1} + \beta_4 \frac{D}{P} + u_t, \quad (22)$$

where y_t denotes the log monthly real return, inf_{t-1} is the inflation rate, variable rrf_{t-1} is the relative T-bill rate constructed using the 3-month T-bill rate from Ibbotson Associates (1999), the difference between the current T-bill rate and the moving average of the previous 12-month rates, and D/P is the dividend–price ratio. We include Model I to show the importance of the inflation rate. Model II has been used by Campbell (1991). Model III includes all the predictors. By no means, these predictors are exhaustive lists of all the discovered predictors. Since the focus of this paper is to propose one way of measuring economic significance of predictability, the reported results are only meant to suggest the economic significance of these variables.

We use the same estimation procedure that was used in Section 3. Table 3 reports results based on the value-weighted NYSE/AMEX/NASDAQ index returns, the S&P 500 index returns⁴ and the equal-weighted NYSE/AMEX/NASDAQ index returns from January 1952 to December 1998.

The most significant result is that our trading strategy, no matter which of the three different models we base it on, improves the average monthly return dramatically and substantially reduces volatilities at the same time. The only exception is for the S&P 500 index returns under Model II. This could due to the fact that we are not using the actual dividend yield for the index due to lacking of data. Despite that, the Sharpe ratio is still much higher for the active trading strategy. In general, using the value-weighted NYSE/AMEX/NASDAQ composite index, we achieve an average log real return of 0.739% for Model I, 0.721% for Model II and 0.823% for Model III, compared to a mean return of 0.663% for the buy-and-hold strategy. The corresponding volatility decreases from 4.19% to 3.52%, 3.05% and 3.00% for the three models, respectively. In other words, the gain in return is as high as 24% and the reduction in volatility is 28%.

⁴ CRSP tape does not report dividends for S&P 500 index. Therefore, we construct the total returns by adding dividends from NYSE/AMEX/NASDAQ composite index.

Table 3
Comparison of different models (1952–1998)

Transactions costs	Buy and hold	Model I			Model II			Model III		
		0 BP	10 BP	20 BP	0 BP	10 BP	20 BP	0 BP	10 BP	20 BP
<i>S&P 500 index returns</i>										
Mean return (%)	0.6826	0.7665	0.7590	0.7516	0.6268	0.6002	0.5736	0.8065	0.7902	0.7739
Volatility (%)	4.1679	3.6672	3.6682	3.6695	3.1883	3.1900	3.1924	3.1993	3.2002	3.2016
R ² (%)		5.9244			3.6039			8.7219		
(S.D.)		(1.783)			(1.113)			(1.097)		
Cumulative return	45.989	74.398	71.297	68.323	33.304	28.526	24.413	93.495	85.189	77.614
(μ/σ)	0.1638	0.2090	0.2069	0.2048	0.1966	0.1882	0.1797	0.2521	0.2469	0.2417
(Sharpe ratio)										
<i>NYSE/AMEX/NASDAQ value weighted index returns</i>										
Mean return (%)	0.6626	0.7392	0.7275	0.7158	0.7209	0.6925	0.6641	0.8225	0.8030	0.7835
Volatility (%)	4.1900	3.5165	3.5167	3.5172	3.0508	3.0526	3.0550	2.9992	2.9994	3.0001
R ² (%)		6.5935			4.4874			9.3112		
(S.D.)		(2.246)			(0.837)			(1.371)		
Cumulative return	40.968	63.661	59.531	55.665	57.303	48.683	41.337	102.44	91.669	82.016
(μ/σ)	0.1581	0.2102	0.2069	0.2035	0.2363	0.2269	0.2174	0.2743	0.2677	0.2612
(Sharpe ratio)										
<i>NYSE/AMEX/NASDAQ equal weighted index returns</i>										
Mean return (%)	0.7733	0.8422	0.8255	0.8089	0.9864	0.9576	0.9289	1.0782	1.0570	1.0357
Volatility (%)	4.9726	4.0749	4.0758	4.0771	3.4705	3.4737	3.4774	3.4760	3.4759	3.4762
R ² (%)		6.9331			5.6318			9.0714		
(S.D.)		(1.779)			(0.831)			(0.921)		
Cumulative return	77.357	114.58	104.21	94.769	259.60	220.63	187.48	436.59	387.11	343.22
(μ/σ)	0.1555	0.2067	0.2025	0.1984	0.2842	0.2757	0.2671	0.3102	0.3041	0.2979
(Sharpe ratio)										

This table compares average returns using different trading strategies. Model I includes lagged return and inflation rate. Model II includes lagged return, dividend–price ratio and relative interest rate. Model III includes all of the predictors used in Models I and II. I use returns data from 1942 to 1951 to start the initial estimation of the model parameters. The active trading strategy requires updating model estimates each month from 1952 to 1998 and making one-period-ahead forecasts. If the forecasted return is greater than the risk-free rate, I invest in stock for the next period. Otherwise, I invest in a risk-free bond that returns R_f next period. All returns are real log returns.

These results also show that the inflation rate is an important predictor, not only for R^2 , but also for the economic benefits. The cumulative gains increase nearly 55% for Model I. In contrast, Model II, which includes the dividend yield and relative interest rate, only gains by about 40% in the cumulative return. When all the variables are included (Model III), the cumulative gain is 1.5 times higher. Quantitatively, the cumulative gains are enormous. For example, under the buy-and-hold strategy, a dollar invested in 1952 becomes 41 dollars in 1998. If we had implemented the active trading strategy with Model III, our return would be

102 dollars. Even under Model II, which relies on an R^2 in the magnitude of 4.5%, we would have gained an additional 16 dollars.

We can also use the Sharpe ratio to evaluate the overall performance. The buy-and-hold strategy has a Sharpe ratio of 0.1581, but the active trading strategy increases the Sharpe ratio by about 73%. Such significant economic benefits of predictability remain when there are transactions costs. For example, when the cost is 0.1% per transaction, under Model III the average monthly return is 0.803% per month. If the transactions costs double, we still see an average monthly return of 0.784% (an 18% increase). The corresponding Sharpe ratios remain substantially higher than those under a buy-and-hold strategy.

We perform a similar exercise on both equal-weighted NYSE/AMEX/NASDAQ and the S&P 500 index returns. These results are also reported in Table 3. The general conclusion remains. Overall, there is an even more impressive performance when using the equalweighted index, but a little smaller gains for the S&P 500 index. For the equal-weighted index, the average returns are 0.842% for Model I, 0.986% for Model II and 1.078% for Model III, compared with 0.773% for the buy-and-hold strategy. The relative gain from the buy-and-hold strategy is as high as 39%. Under Model III, the cumulative return is more than five times higher and the return volatility is much lower than that from the buy-and-hold strategy. Therefore, the Sharpe ratio is still as high as 0.298 after taking into account transactions costs of 0.2%. This almost doubles that of the buy-and-hold strategy. Unlike in the case of S&P 500 index returns, the dividend–price ratio and the relative interest rate are important variables too.

The performance for the S&P 500 index seems a little higher (0.767%) for Model I and a little weaker (0.807%) for Model III than the corresponding results using the value-weighted index. As mentioned before, the results are worse for Model II since we are approximating the dividend yield using that from the value-weighted NYSE/AMEX/NASDAQ index. Returns from the trading strategy also exhibit much lower volatilities. For example, the volatilities are 12%, 24% and 23% lower, respectively, than that of a buy-and-hold strategy for the three models. The Sharpe ratios are large and vary from 0.197 to 0.252 under different models, which corresponds to a 20% increase even for the worst case. In the presence of transactions costs, the trading strategy can still substantially improve performance relative to the buy-and-hold strategy.

At the same time, we do not suggest that the active strategy is superior to a buy-and-hold strategy in each single testing period. In order to be informative, we also report the average relative monthly returns with respect to the buy-and-hold returns for the active strategy implemented based on Model III over each 5-year period in Table 4.⁵ For the S&P 500 index, the trading strategy has returned much higher than the buy-and-hold strategy 7 out of the 10 subsample periods. In most of the subsample periods, returns are substantial. Results are similar for value-weighted NYSE/AMEX/NASDAQ index. However, for the equal-weighted NYSE/AMEX/NASDAQ index, results are stronger as usual. Not only during 8 out of the 10 subsample periods, we observe better performance for the active

⁵ The last subsample period consists only 2 years.

Table 4

Five-year performance comparison for different index returns (1952–1998)

	Buy and hold (%)	S&P 500 index			VW index			EW index		
		0 BP	10 BP	20 BP	0 BP	10 BP	20 BP	0 BP	10 BP	20 BP
1952–1956	0.341	0.000	0.000	0.000	−0.126	−0.136	−0.146	−0.002	−0.033	−0.036
1957–1961	0.915	−0.101	−0.111	−0.121	−0.028	−0.045	−0.061	0.041	0.028	0.014
1962–1966	0.310	0.218	0.188	0.158	0.381	0.355	0.328	0.455	0.429	0.402
1967–1971	0.322	0.259	0.242	0.226	0.228	0.201	0.175	0.417	0.392	0.367
1972–1976	−0.215	0.826	0.810	0.793	0.816	0.803	0.790	1.269	1.244	1.219
1977–1981	−0.033	0.210	0.197	0.183	0.265	0.242	0.219	0.004	−0.026	−0.056
1982–1986	0.208	0.069	0.059	0.049	0.075	0.062	0.048	0.023	0.004	−0.016
1987–1991	0.754	0.353	0.326	0.299	0.594	0.571	0.547	0.622	0.602	0.582
1992–1996	0.898	−0.585	−0.600	−0.615	−0.523	−0.535	−0.546	−0.362	−0.385	−0.408
1997–1998	0.816	−0.218	−0.255	−0.293	−0.452	−0.498	−0.544	1.060	1.026	0.993

This table compares average monthly excess log returns relative to a buy-and-hold returns for different indices over five year period. The active trading strategy is applied to Model III, which includes lagged return, dividend–price ratio, relative interest rate and inflation rate. I use returns data from 1942 to 1951 to start the initial estimation of the model parameters. The active trading strategy requires updating model estimates each month from 1952 to 1998 and making one-period ahead forecasts. If the forecasted return is greater than the risk-free rate, I invest in stock for the next period. Otherwise, I invest in a risk-free bond that returns R_f next period. All returns are real log returns.

strategy, but the average return is more than doubled in the most recent period as well. Therefore, the results are robust to sample periods.

4.3. Bootstrapping the actual returns

The observed historical returns are just one draw from an unobservable stochastic return process. To study the reliability of measuring the economic benefit of small predictability through active trading strategy, we design the following bootstrapping method. For a given model, we first estimate residuals from,

$$y_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \beta X_t + e_t, \quad (23)$$

where X_t includes other predictors. Then we randomly draw T observations with replacement $(e_1^s, \dots, e_T^s)$ from the estimated residuals $(e_1, \dots, e_T)$, i.e., $e_t^s = e_{s_t}$. We constructed returns (y_t^c) according to

$$y_t^c = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1}^c + \beta X_{s_t} + e_t^s. \quad (24)$$

With these returns generated from an empirical distribution, we compute average returns from the active trading strategy based on the three proposed models. We repeat this process 1000 times. Table 5 reports the average values with and without transactions costs for different models and for indexes.

For S&P 500 index returns, there is a 24% increase in the average monthly return for Model III over the buy-and-hold return. Such an increase is much higher than the one reported in Table 3. Model II also produces higher average return now. Increases under

Table 5

Comparison of different models by using simulated returns (1952–1998)

Transactions costs	Buy and hold	Model I			Model II			Model III		
		0 BP	10 BP	20 BP	0 BP	10 BP	20 BP	0 BP	10 BP	20 BP
<i>S&P 500 index returns</i>										
Mean return (%)	0.6718	0.7664	0.7435	0.7206	0.7282	0.6910	0.6539	0.8322	0.7929	0.7535
Return difference		0.0946	0.0717	0.0488	0.0565	0.0193	−0.0179	0.1605	0.1211	0.0818
(S.D.)		(0.0027)	(0.0026)	(0.0025)	(0.0029)	(0.0028)	(0.0027)	(0.0031)	(0.0030)	(0.0029)
Better than “buy and hold” (%)		85.800	81.400	73.700	73.100	53.700	37.600	96.200	90.100	80.800
Volatility (%)	4.1725	3.7326	3.7341	3.7360	3.3766	3.3780	3.3801	3.3583	3.3604	3.3632
R ² (%)		5.4530			3.7079			8.4936		
Cumulative return	72.692	107.04	95.498	85.280	89.675	74.089	61.337	150.21	122.25	99.638
(μ/σ) (Sharpe ratio)	0.1620	0.2058	0.1995	0.1932	0.2155	0.2042	0.1928	0.2478	0.2357	0.2237
<i>NYSE/AMEX/NASDAQ value weighted index returns</i>										
Mean return (%)	0.6530	0.7375	0.7140	0.6906	0.7239	0.6858	0.6477	0.8146	0.7759	0.7372
Return difference		0.0845	0.0611	0.0376	0.0710	0.0329	−0.0052	0.1617	0.1230	0.0842
(S.D.)		(0.0028)	(0.0026)	(0.0025)	(0.0031)	(0.0030)	(0.0029)	(0.0033)	(0.0032)	(0.0031)
Better than “buy and hold” (%)		83.300	77.400	66.900	76.300	62.000	43.400	94.800	90.000	80.800
Volatility (%)	4.1824	3.7011	3.7025	3.7045	3.3373	3.3388	3.3410	3.3228	3.3248	3.3275
R ² (%)		5.8529			4.3457			8.8427		
Cumulative return	66.713	91.836	81.860	73.037	86.092	70.579	57.945	137.19	112.08	91.685
(μ/σ) (Sharpe ratio)	0.1571	0.1997	0.1932	0.1866	0.2168	0.2051	0.1933	0.2451	0.2332	0.2212
<i>NYSE/AMEX/NASDAQ equal weighted index returns</i>										
Mean return (%)	0.7656	0.8642	0.8417	0.8192	0.8983	0.8607	0.8231	0.9487	0.9128	0.8768
Return difference		0.0986	0.0761	0.0537	0.1327	0.0951	0.0574	0.1831	0.1471	0.1112
(S.D.)		(0.0034)	(0.0033)	(0.0032)	(0.0040)	(0.0039)	(0.0038)	(0.0043)	(0.0042)	(0.0041)
Better than “buy and hold” (%)		83.300	76.000	67.200	85.200	76.100	65.200	92.300	86.600	79.500
Volatility (%)	4.9468	4.3613	4.3628	4.3647	3.9127	3.9148	3.9174	3.9040	3.9063	3.9091
R ² (%)		6.1862			5.7437			8.7136		
Cumulative return	189.12	254.21	230.47	209.19	287.55	237.95	197.20	354.36	295.47	246.73
(μ/σ) (Sharpe ratio)	0.1558	0.1985	0.1932	0.1879	0.2295	0.2196	0.2097	0.2431	0.2336	0.2241

This table compares average returns by using different trading strategies. The returns are generated from the actual return residuals of different models. Model I includes lagged return and inflation rate. Model II includes lagged return, dividend–price ratio and relative interest rate. Model III includes all of the predictors used in Models I and II. I use returns data from 1942 to 1951 to start the initial estimation of the model parameters. The active trading strategy requires updating model estimates each month from 1952 to 1998 and making one-period-ahead forecasts. If the forecasted return is positive, I invest in stock for the next period. Otherwise, I invest in a risk-free bond that returns R_f next period. All returns are real log returns.

Model I are closer to those reported in Table 3. Since decreases in the return volatilities are less than those reported in Table 3, the Sharpe ratios are comparable.

We find it interesting that the trading strategy under Model III outperforms the buy-and-hold strategy 96% of times. Even with transactions costs of 20 bases points, the strategy still outperforms 81% of times. In contrast, when there are large transactions costs, Model II underperforms the buy-and-hold strategy as before. Therefore, Models I and III are more reliable.

When we use value-weighted NYSE/AMEX/NASDAQ index under both Models I and III, results are very similar to those observed in Table 3. Again, when there are large transactions costs, the active trading strategy still outperforms the buy-and-hold strategy 81% of times under Model III and 67% of times under Model I. Again, Model II does not show superior performance when there are large transactions costs. Despite that, the Sharpe ratio is still 23% higher than that of the buy-and-hold returns. Although the performance is not as impressive as that reported in Table 3 when the equal-weighted NYSE/AMEX/NASDAQ index is used, the trading strategy's average monthly returns under Models I, II and III are still 13%, 17% and 24% higher than the buy-and-hold returns. It is also interesting that Model II is better than Model I this time no matter how we evaluate. Even with 0.2% transactions costs under Model II, the active strategy outperforms a buy-and-hold strategy 65% of times.

These analyses demonstrate the economic significance of small levels of predictability. This conclusion is also reliable because we have used both actual returns as well as simulated returns based on parametric distributions and observed empirical distributions. The sizable economic profits generated based on this information is also robust to transactions costs.

5. Conclusions

Many researchers have focused on detecting predictability in stock market returns. In this study, we examine the issue from the perspective of measuring economic significance of known predictability. Because it is difficult to argue the importance of predictability in terms of an R^2 in the magnitude of 2% from a statistical point of view, we measure the economic significance through the performance of a market timing strategy that uses a small R^2 . We demonstrate that when such a trading strategy is applied to historical data for the postwar period, aggregate market return can be doubled. Through simulations and bootstrapping on realized returns, we also show that such a strategy can be used reliably to measure the economic significance of return predictability. Our results ensure our confidence in the economic significance of weak predictability. We have also carried out theoretical analysis based on Student- t distributed stock returns, which allows for analyzing the impact of kurtosis.

In light of our findings with regard to economic benefits of small levels of predictability, what are the implications to market efficiency? We believe that past markets can still be conditionally efficient in the following sense: If the trading strategies studied here were not in the information set of past investors, we should not alter our beliefs about past markets. When there are market imperfections, such as liquidity

traders, informed investors can profit by taking advantage of different groups of noisy traders. That is, weak serial correlation compensates investors for added noisy supply risk. Finally, market efficiency arguments inherit a self-destructive force. If there is no rational reason for predictability, it is hard to perceive whether such phenomena will persist when every investor tries to use this trading strategy. In fact, predictability has weakened in recent years.

Acknowledgements

I am grateful to John Y. Campbell, Ted Day, Chris Kirby, Burton G. Malkiel, Larry Merville, Franz C. Palm (the editor), the anonymous referee and seminar participants at the 2000 FMA annual meeting for their comments.

Appendix A. Proof of Proposition 1

To prove Proposition 1, we first to establish the following lemma:

Lemma 1. *For any positive scale λ , the following inequality holds as long as $\beta \geq ((v-1)/v)(\lambda^2/(1+\lambda^2))$:*

$$\Psi_v(\lambda) + \frac{\beta}{v-1} \frac{v + \lambda^2}{\lambda} \psi_v(\lambda) > 1, \quad (\text{A.1})$$

where $\psi_v(\lambda) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}} \left[1 + \frac{\lambda^2}{v}\right]^{-\frac{v+1}{2}}$ and $\Psi_v(\lambda) = \int_{-\infty}^{\lambda} \psi_v(w)dw$.

Proof. Let $g(x) = \int_{-\infty}^x \left(1 + \frac{w^2}{v}\right)^{-\frac{v+1}{2}} dw + \frac{v\beta}{(v-1)x} \left(1 + \frac{x^2}{v}\right)^{-\frac{v-1}{2}}$. Since $g'(x) = \left[1 - \beta \frac{v(1+x^2)}{(v-1)x^2}\right] \left(1 + \frac{x^2}{v}\right)^{-(v+1)/2} < 0$ when $\beta \geq (v-1)x^2/[v(1+x^2)]$, $g(x)$ is monotonically decreasing in x . Moreover, there is a limiting point with $(\Gamma(v+1/2)/[\Gamma(v/2)\sqrt{v\pi}])g(+\infty) \rightarrow 1$. The inequality is established. $\square$

When y_t and $\hat{y}_t$ are jointly distributed as Student- t with v degree of freedom as in Eq. (1), we have the following conditional moments according to Zellner (1971),

$$E[y_t | \hat{y}_t] = \mu + \rho \frac{\sigma_1}{\sigma_2} (\hat{y}_t - \mu) \quad (\text{A.2})$$

$$\text{Var}[y_t | \hat{y}_t] = \frac{v(1-\rho^2)\sigma_1^2}{v-1} \left[1 + \frac{1}{v\sigma_2^2} (\hat{y}_t - \mu)^2\right] \quad (\text{A.3})$$

Therefore, the expected return under $\mathcal{F}$ can be written as

$$m = E[y_t | \mathcal{F}] = \text{Prob}(\hat{y}_t > 0)E[y_t | \hat{y}_t > 0] + \text{Prob}(\hat{y}_t \leq 0)E[0 | \hat{y}_t \leq 0] \quad (\text{A.4})$$

$$\begin{aligned} &= \text{Prob}(\hat{y}_t > 0)E[E(y_t | \hat{y}_t) | \hat{y}_t > 0] \\ &= \int_0^\infty \left[\mu + \rho \frac{\sigma_1}{\sigma_2} (\hat{y}_t - \mu) \right] \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}\sigma_2^2} \left[1 + \frac{1}{v\sigma_2^2} (\hat{y}_t - \mu^2) \right]^{-\frac{v+1}{2}} d\hat{y}_t \\ &= \mu \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}} \left[\int_{-\infty}^{\lambda} \left(1 + \frac{w^2}{v} \right)^{-\frac{v+1}{2}} dw + \frac{v\beta}{\lambda(v-1)} \left(1 + \frac{\lambda^2}{v} \right)^{-\frac{v+1}{2}} \right] \\ &= \mu \left[\Psi_v(\lambda) + \frac{\beta}{v-1} \frac{v + \lambda^2}{\lambda} \psi_v(\lambda) \right]. \end{aligned} \quad (\text{A.5})$$

In the above derivation, the third equality is due to the law of iterative expectation; and the fourth equality is from Eq. (A.2). In the fifth equality, we have defined $\lambda = (\mu/\sigma_2)$, $w = (\hat{y}_t - \mu)/(\sigma_2)$ and $\beta = (\rho\sigma_1)/(\sigma_2)$.

When $\beta \geq \frac{(v-1)\lambda^2}{v(1+\lambda^2)} = \frac{(v-1)\mu^2}{v(\mu^2 + \sigma_2^2)}$ by applying the Lemma 1, we show that $m > \mu$. In the special case when $\rho\sigma_1 = \sigma_2$, the proposition condition is naturally satisfied since $\beta = 1$. Similarly, before proving the relation on volatilities, we use the following result:

Lemma 2. For any positive scale λ and a scale $|\rho| \leq 1$, the following inequality holds as long as $\beta \geq \frac{(v-1)\lambda^2}{v(1+\lambda^2)}$.

$$\left[\lambda^2 + \frac{v\beta^2}{(v-2)\rho^2} \right] [1 - \Psi_v(\lambda)] + \frac{v\beta^2}{v-2} [\Psi_v(\lambda) - \Psi_{v-2}(\lambda)] \geq (2\beta - \beta^2) \frac{v\lambda + \lambda^3}{v-1} \psi_v(\lambda). \quad (\text{A.6})$$

Proof. Let $g(x) = \left[x^2 + \frac{v\beta^2}{(v-2)\rho^2} \right] [\Psi_v(x) - 1] + \frac{v\beta^2}{v-2} [\Psi_{v-2}(x) - \Psi_v(x)] + (2\beta - \beta^2) \frac{vx + x^3}{v-1} \psi_v(x)$. Since

$$\begin{aligned} g'(x) &= 2x[\Psi_v(x) - 1] + \left[x^2 + \frac{v\beta^2}{(v-2)\rho^2} \right] \psi_v(x) - \frac{v\beta^2}{v-2} \psi_v(x) + \frac{\beta^2}{v-1} (v + x^2) \\ &\quad \times \psi_v(x) + \frac{\beta(2-\beta)}{v-1} (v + 3x^2) \psi_v(x) - \frac{\beta(2-\beta)}{v-1} (v+1)x^2 \psi_v(x) \\ &\geq -2\beta \frac{v+x^2}{v-1} \psi_v(x) + x^2 \psi_v(x) + \frac{v\beta^2}{v-2} (1/\rho^2 - 1) \psi_v(x) + \frac{\beta^2}{v-1} (v \\ &\quad + x^2) \psi_v(x) + \frac{\beta(2-\beta)}{v-1} (v + 2x^2 - vx^2) \psi_v(x) \\ &= (1-\beta)^2 x^2 \psi_v(x) + (1-\rho^2) \frac{v\beta^2}{(v-2)\rho^2} \psi_v(x) > 0 \end{aligned} \quad (\text{A.7})$$

The first inequality uses Lemma 1. Therefore, $g(x)$ is monotonically increasing in x , with a limiting value of $g(+\infty) \rightarrow 0$. This implies that $g(x)$ is a negative function. $\square$

The volatility of the trading strategy can then be shown as

$$\begin{aligned}
 \text{Var}[y_t | \mathcal{F}] &= E[y_t^2 | \mathcal{F}] - m^2 = \text{Prob}(\hat{y}_t > 0)E[y_t^2 | \hat{y}_t > 0] + \text{Prob}(\hat{y}_t \leq 0)E[0 | \hat{y}_t \leq 0] - m^2 \\
 &= \text{Prob}(\hat{y}_t > 0)E[E^2(y_t | \hat{y}_t) + E_{\hat{y}_t}[\text{Var}(y_t | \hat{y}_t)] | \hat{y}_t > 0] - m^2 \\
 &\leq \int_0^\infty \left\{ \left[\mu + \rho \frac{\sigma_1}{\sigma_2} (\hat{y}_t - \mu) \right]^2 + \frac{1 - \rho^2}{v - 2} v \sigma_1^2 \right\} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi} \sigma_2^2} \\
 &\quad \times \left[1 + \frac{1}{v\sigma_1^2} (\hat{y}_t - \mu)^2 \right]^{-\frac{v+1}{2}} d\hat{y}_t - \mu^2 \\
 &= \left[\mu^2 + \frac{1 - \rho^2}{v - 2} v \sigma_1^2 \right] \Psi_v(\lambda) + \int_{-\infty}^{\lambda} (2\mu\rho\sigma_1 x + \rho^2 \sigma_1^2 x^2) \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} \\
 &\quad \times \left[1 + \frac{1}{v\sigma_1^2} x^2 \right]^{-\frac{v+1}{2}} dx - \mu^2 \\
 &= \left[\mu^2 + \frac{1 - \rho^2}{v - 2} v \sigma_1^2 \right] \Psi_v(\lambda) + \left[\frac{2\mu\rho\sigma_1}{v - 1} - \frac{\rho^2 \sigma_1^2 \lambda}{v - 1} \right] (v + \lambda^2) \psi_v(\lambda) \\
 &\quad + \frac{v\rho^2 \sigma_1^2}{v - 2} \Psi_{v-2}(\lambda) - \mu^2 \\
 &= \frac{v\sigma_1^2}{v - 2} + \sigma_2^2 \left[(\lambda^2 + v \frac{\sigma_1^2}{\sigma_2^2}) [\Psi_v(\lambda) - 1] + \frac{v\beta^2}{v - 2} [\Psi_{v-2}(\lambda) - \Psi_v(\lambda)] \right. \\
 &\quad \left. + (2\beta - \beta^2) \frac{v\lambda + \lambda^3}{v - 1} \psi_v(\lambda) \right] \tag{A.8}
 \end{aligned}$$

In the second equality, we have applied the rule of variance decomposition. The first inequality uses the result of $\mu < m$. The fourth equality uses integration by parts. Again, using Lemma 2, we show the following result:

$$\text{Var}[y_t | \mathcal{F}] < \text{Var}[y_t]. \tag{A.9}$$

Appendix B. Variances of estimators

The following results are standard in a time-series analysis:

$$\begin{cases} \phi_0 = \mu(1 - \phi_1 - \phi_2), \\ \gamma_0 = \text{Var}(y_t) = \frac{v}{v - 2} \sigma_1^2 = \frac{\sigma^2}{(1 - \phi_2^2)(1 - \theta^2)}, \\ \gamma_1 = \text{Cov}[y_t, y_{t-1}] = \theta\gamma_0 \end{cases} \tag{B.1}$$

where $\theta = \phi_1 / (1 - \phi_2)$. Given the DGP of Eq. (6) and under the law of large numbers, the residual from an AR(1) model is

$$u_t = y_t - (\hat{\phi}_0 + \hat{\phi}_1 y_{t-1}) \rightarrow (\phi_1 - \theta) \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + e_t, \quad (\text{B.2})$$

where $\tilde{y}_t$ is the demeaned y_t . From there, we can compute the residual variance as $\sigma_u^2 = \text{Var}(u_t) = \lambda_0(1 - \theta^2)$. Therefore, we can derive the variance–covariance matrix as,

$$\text{Var} \left(\begin{bmatrix} \hat{\phi}_0 \\ \hat{\phi}_1 \end{bmatrix} \right) = \frac{\sigma_u^2}{T} \begin{bmatrix} 1 & E y_t \\ E y_t & E y_t^2 \end{bmatrix}^{-1} = \frac{1 - \theta^2}{T} \begin{bmatrix} \gamma_0 + \mu^2 & -\mu \\ -\mu & 1 \end{bmatrix}. \quad (\text{B.3})$$

When the estimation model (an AR(2) in this case) is the same as the DGP, it is straightforward to show the following variances and covariances of estimators based on Eq. (B.1):

$$\begin{aligned} \text{Var} \left(\begin{bmatrix} \hat{\phi}_0 \\ \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} \right) &= \frac{\sigma^2}{T} \begin{bmatrix} 1 & E y_t & E y_t \\ E y_t & E y_t^2 & E y_t y_{t-1} \\ E y_t & E y_t y_{t-1} & E y_t^2 \end{bmatrix}^{-1} \\ &= \frac{\sigma^2}{T} \begin{bmatrix} 1 + \frac{2\mu_2}{\gamma_0 + \gamma_1} & -\frac{\mu}{\gamma_0 + \gamma_1} & -\frac{\mu}{\gamma_0 + \gamma_1} \\ -\frac{\gamma_0}{\gamma_0^2 - \gamma_1^2} & -\frac{\gamma_1}{\gamma_0^2 - \gamma_1^2} & \\ & & \frac{\gamma_0}{\gamma_0^2 - \gamma_1^2} \end{bmatrix}. \end{aligned} \quad (\text{B.4})$$

We use these results to obtain Eqs. (12) through (19).

References

- Barberis, N., 2000. Investing for the long run when returns are predictable. *Journal of Finance* 54, 225–264.
- Boudoukh, J., Richardson, M.P., Whitelaw, R.F., 1994. A tale of three schools: insights on autocorrelations of short-horizon security returns. *Review of Financial Studies* 7, 539–573.
- Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance* 47, 1731–1764.
- Campbell, J.Y., 1987. Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.
- Campbell, J.Y., 1991. A variance decomposition for stock returns. *Economic Journal* 101, 157–179.
- Campbell, J.Y., Shiller, R.J., 1988. The dividend–price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195–228.
- Cochrane, J.H., 1999. Portfolio advice for a multifactor world. *Economic perspectives*. Federal Reserve Bank of Chicago 23, 59–77.
- Conrad, J., Kaul, G., 1988. Time-variation in expected returns. *Journal of Business* 61, 409–425.

- Conrad, J., Kaul, G., 1989. Mean reversion in short-horizon expected returns. *Review of Financial Studies* 2, 225–240.
- Cutler, D.M., Poterba, J.M., Summers, L.H., 1991. Speculative dynamics. *Review of Economic Studies* 58, 529–546.
- Elliott, G., Timmermann, A., 2002. Optimal forecast combinations under general loss functions and forecast error distributions. Working Paper, UC-San Diego.
- Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25, 383–417.
- Fama, E.F., 1991. Efficient capital markets: II. *Journal of Finance* 46, 1575–1617.
- Fama, E.F., French, K.R., 1988a. Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246–273.
- Fama, E.F., French, K.R., 1988b. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–27.
- Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23–49.
- Fisher, L., 1966. Some new stock-market indexes. *Journal of Business* 39, 191–225.
- Granger, C.W.J., Pesaran, M.H., 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19, 537–560.
- Greene, W.H., 2000. *Econometric Analysis*. Prentice-Hall, Upper Saddle River, p. 899.
- Hirshleifer, J., 1975. Speculation and equilibrium: information, risk, and markets. *Quarterly Journal of Economics* 89, 519–542.
- Ibbotson Associates, 1999. *Stocks, bonds, bills, and inflation. 1999 Yearbook*. Chicago.
- Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset allocation perspective. *Journal of Finance* 51, 385–424.
- Kaul, G., 1996. Predictable components in stock returns. In: Maddala, G.S., Rao, C.R. (Eds.), *Handbook of Statistics Series: Statistical Methods of Finance*. Elsevier, Oxford, pp. 67–83.
- Kaul, G., Nimalendran, M., 1990. Price reversals: bid-ask errors or market overreaction? *Journal of Financial Economics* 28, 67–83.
- Keim, D., Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357–390.
- Kirby, C., 1997. Measuring the predictable variation in stock and bond returns. *Review of Financial Studies* 10, 579–630.
- LeRoy, S.F., 1973. Risk aversion and the martingale property of stock returns. *International Economic Review* 14, 436–446.
- Lo, A.W., MacKinlay, A.C., 1990a. When are contrarian profits due to market overreaction? *Review of Financial Studies* 3, 175–205.
- Lo, A.W., MacKinlay, A.C., 1990b. An econometric analysis of nonsynchronous trading. *Journal of Econometrics* 45, 181–211.
- Lucas, R.E., 1978. Asset prices in an exchange economy. *Econometrica* 46, 1429–1446.
- Maddala, G.S., 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge Univ. Press, Cambridge, pp. 165–170.
- McDonald, J.B., Xu, Y., 1995. A generalization of the beta distribution with applications. *Journal of Econometrics* 66, 133–152.
- Pesaran, M.H., Timmermann, A., 1995. The robustness and economic significance of predictability of stock returns. *Journal of Finance* 50, 1201–1228.
- Poterba, J., Summers, L.H., 1988. Mean reversion in stock returns: evidence and implications. *Journal of Financial Economics* 22, 27–60.
- Shiller, R.J., 1981. Do stock prices move too much to be justified by subsequent changes in dividend. *American Economic Review* 71, 421–436.
- Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York, pp. 383–389.