Are Growth and Value More Predictable than the Market?

A Model Selection Approach

Xiaojing Su
Department of Economics
Texas A&M University
College Station, TX 77843, USA

Jian Yang*
The Business School
PO Box 173364
University of Colorado Denver
Denver, CO 80217-3364

This Version: September 2007

We thank Yongmiao Hong for sharing some computer programs used in this study, and Dennis W. Jansen, Qi Li, Hui Guo, Donald R. Fraser and seminar participants at Texas A&M University for much help and suggestions.

*Corresponding author. Email: Jian.Yang@cudenver.edu. Tel: (303) 556-5852; Fax: (303) 556-5899

Are Growth and Value More Predictable than the Market?

A Model Selection Approach

Abstract

This paper employs a model selection approach to investigate the out-of-sample international stock return predictability based on past price information. In particular, we use several nonparametric nonlinear models to address possible nonlinearity-in-mean predictability; we also adopt economic criteria, in addition to commonly used statistical criteria, to evaluate the forecasting performance. For thirteen major international stock markets, growth stocks appear to be more predictable than the general stock markets and value stocks, especially when evaluated with economic criteria.

Key words: international stock markets; model selection; economic criteria; nonparametric models; forecasting.

JEL Classification: C2, C5, F3

Are Growth and Value More Predictable than the Market?

A Model Selection Approach

1. Introduction

Whether stock returns are predictable or not has been one of the most active research topics in finance, especially since Fama's (1965) seminal work on the efficient market hypothesis. In particular, numerous authors have investigated the weak-form market efficiency, which implies that security prices are unpredictable based on historical price information. Most financial economists would agree that stock return predictability is crucial for understanding the dynamics of financial markets, although there is an ongoing debate about whether it implies market inefficiency (e.g., Fama, 1965) or is consistent with rational pricing (e.g., Campbell and Cochrane, 1999). The inference on stock predictability also has implications for practitioners. For example, portfolio managers can use it to develop trading strategies, and financial managers can use it for the equity financing decision.

However, most relevant to this study, the majority of earlier works on testing the predictability of common stock returns based on *past returns* fails to reject the random walk hypothesis. One possible explanation is the lack of power in empirical techniques (e.g., the autocorrelation test) used in these studies. To address this issue, Lo and MacKinlay (1988) propose the now-commonly-used variance ratio test, but the new test at best yields a mixed conclusion, especially for the low-frequency, e.g., monthly, data (e.g., Ayadi and Pyun, 1994; Urrutia, 1995; Coggin, 1998; Abraham et al., 2002; Chaudhuri and Wu, 2003; and Patro and Wu, 2004). The lack of compelling empirical evidence leads both Malkiel (2005) and Chordia, Roll, and Subrahmanyam (2005) to conclude that asset prices are difficult to predict for most financial economists and professionals. This conclusion also appears to apply to the empirical test of the semi-strong form market efficiency (for example, exploring the predictive power of various state variables for stock returns), which we do not address in this paper. In particular, while many authors show that some financial variables help forecast the equity premium, Goyal and Welch (2006) comprehensively examine the extant empirical evidence using the data updated to 2004 but find little support for stock return predictability, especially in the out-of-sample test.

Yet, there is another alternative explanation that warrants further investigation. Both the popular autocorrelation test and the variance ratio test assume linearity and only test serial uncorrelatedness rather than martingale difference (Hsieh, 1991; McQueen and Thorley, 1991; Hong, 1999; and Hong and Lee, 2003). However, a nonlinear time series can have zero autocorrelation but a non-zero mean conditional on its past history (i.e. predictable based on the past history). That is, both tests may fail to capture predictable nonlinearities in mean and could yield misleading conclusions in favor of the martingale (or loosely random walk) hypothesis.[1] In this paper, we reexamine international stock return predictability by using a comprehensive set of nonlinear models.

In particular, our study contributes to the literature along two important dimensions. First, while most authors focus on stock market indexes, we also comprehensively examine the martingale behavior of two most important equity style indexes, i.e., growth and value stock indexes, for thirteen major international stock markets over the period January 1975 to December 2004.[2] Testing martingale behavior of style equity indexes is important *in itself* because style investing has become popular in last two decades and its performance benchmarks are style

[1] The terms "random walk" and "martingale" have been interchangeably used in the efficient capital markets literature. However, as discussed in Fama (1965), it is the martingale property (or unpredictability) of security prices that is of essential interest to this huge body of the literature. Also see Granger (1992) for a more recent elaboration on this point. Strictly speaking, the innovations series is independent and identically distributed for "random walk", while it is a martingale difference sequence for "martingale." Nevertheless, testing the martingale property is still conventionally termed "random walk hypothesis" (see, e.g., Granger, 1992), which is also followed in this study (wherever appropriate).
[2] The major equity style categories are (1) value and growth and (2) small and large. However, and the style investing of small and large receives much less attention than that of value and growth possibly because many authors, e.g., Hogan et al. (2004), find that the size premium associated with the style investing of small and large has substantially attenuated since the 1980s.

indexes (see, e.g., Barberis and Shleifer, 2003). More importantly, equity style indexes can potentially enhance the power of our tests because recent authors suggest that they might be more predictable than general stock market indexes in both behavioral (e.g., Barberis and Shleifer, 2003) and rational pricing (e.g., Campbell and Vuolteenaho, 2004) models.[3] Surprisingly, with the notable exception of Coggin (1998), few authors have addressed the predictability of growth and value portfolios based on their past prices. In this paper, we try to fill this gap by providing a comprehensive investigation of the predictability of style indexes.

Second, we employ the model selection approach (Swanson and White, 1995; 1997) and comprehensively address stock return predictability in a nonlinear, out-of-sample context. Compared to the traditional hypothesis testing approach which the variance ratio test follows, the model selection approach has two advantages. First, it allows us to focus directly on the issue of predictability at hand: out-of-sample forecasting performance, Evaluating out-of-sample forecasting performance is particularly important to making the correct inference based on nonparametric models, as they notoriously tend to overfit the data.[4] Second, unlike the traditional hypothesis testing approach, it does not require the specification of a correct model for its valid application. Our specification also addresses two major methodological deficiencies in this line of literature identified by Granger (1992), who suggests that *benefits can arise…especially from considering non-linear models* and *that only out-of-sample evaluation is relevant and, to some extent, avoids these difficulties (due to data mining).*[5]

[3] Regarding potential nonlinear predictability, Urrutia et al. (2002) also argue that the weak findings of nonlinearities reported in previous research may be primarily attributed to the use of aggregate market index data that can hide nonlinearities at the micro level. They use an (insurance) industry stock index data in their study.
[4] This distinguishes our study from some recent studies (e.g., Al-Khazali, Ding, and Pyun, 2007) using the nonparametric variance ratio test of Wright (2000). Our study also further differs in using several (rather than one) nonparametric nonlinear techniques and exploring economic criteria (which is crucial to the main finding of this study) (e.g., Darrat and Zhong, 2000).
[5] Nonlinear patterns in stock returns might arise because of fads or rational speculative bubbles (McQueen and Thorley, 1991). Hsieh (1991) also suggests that, if the financial market is governed by a not-too-complex chaotic process, it should have short-term nonlinear predictability but not linear predictability. Also see Urrutia et al. (2002) for more related discussions.

A few earlier works (e.g., Patro and Wu, 2004) have addressed the predictability of international stock markets using the variance ratio test. However, these authors only focus on in-sample evidence and, more importantly, fail to allow for potential predicable nonlinearity-in-mean.[6] Also noteworthy, many researchers (Leitch and Tanner, 1991; Brock et al., 1992; Granger, 1992; Gencay, 1998, 1999) have emphasized the importance of trading rule profitability (as an economic criterion) to evaluate the forecasting performance. However, the direction of changes as an alternative economic criterion has not yet been much explored. From a perspective of decision-making under uncertainty, there exist important circumstances under which this criterion is exactly the right one for maximizing the economic welfare of the forecaster (Leitch and Tanner, 1991; Hong and Lee, 2003; Hong and Chung, 2003). Directional predictability in asset returns also has important implications for market timing and the resulting active asset allocation management. In the context of return forecasting based on past price information, we are the first to comprehensively report evidence on both (out-of-sample) trading rule profitability (particularly based on nonlinear models) and the predictability of direction of changes for a number of international stock markets and their stock style indexes.

As conjectured, we find significant out-of-sample predictability in growth style indexes for nine of thirteen countries considered in this paper; however, the evidence is noticeably weaker for value style indexes (five countries) and market indexes (four countries).[7] The main finding that growth stocks are more predictable than value stocks is also quite robust. The remainder of this paper is organized as follows. We discuss econometric methodology in Section 2 and present empirical results in Section 3. We offer some concluding remarks in Section 4.

[6] Note that there is a debate about whether there exists predictable nonlinearity-in-mean in stock prices. For example, although Hsieh (1991) finds little nonlinearity-in-mean in US stock market prices, Gencay (1999) reports nonlinear out-of-sample predictability for similar indexes.

[7] Like many earlier studies, a caveat here is that the inference should be interpreted in light of the limited number of models we examine in this study. In general, martingale means the existence of neither linear nor nonlinear dependence, and we have to test all possible nonlinear dependence to rule out the martingale property of stock returns, which is practically impossible.

2. Econometric Methodology

To forecast stock returns (Y_t) using past returns, we use various models for $E(Y_t \mid I_{t-1})$, where $I_{t-1} = \{Y_{t-1}, Y_{t-2}, ..., Y_{t-d}\}$ is the information set available at time t-1 (where d = 1 for the monthly data used in this study). It is generally believed that Y_t may not be a martingale process and may have dependence in higher moment, and its conditional mean, $E(Y_t \mid I_{t-1})$ is time-varying in a complicated form. We will include various popular nonlinear parametric and nonparametric models used in the literature.

We use the martingale model (with a drift) as benchmark model, and consider following four popular nonlinear models to compare with the benchmark: polynomial regression model (PN), artificial neural network (NN)), functional coefficient model (FC), and nonparametric regression model (NP). We also consider a basic linear model, the autoregression (i.e., AR (1)) model, which is a popular null model for modeling stock returns (Brock et al., 1992).[8] Since the estimation of AR and PN models is relatively simple using the ordinary least squares method, we only briefly discuss below how to implement more complicated econometric tools used in this study (i.e., NN, FC and NP).

2.1 The Artificial Neural Network

Artificial neural networks have been popular in capturing potential nonlinearity-in-mean in financial time series. One big advantage of neural networks over other commonly-used nonlinear time series models is that a class of multilayer neural networks can well approximate a large class of functions. There are usually two types of neural networks -- namely, feedforward and recurrent networks, and we use feedforward networks in this study. The basic structure of

---

[8] We do not consider popular GARCH models in this study. As pointed out in Hsieh (1991), GARCH models capture potential nonlinearity-in-variance, but they are not designed to address nonlinearity-in-mean, which is of major interest in this study. Also see Hong and Lee (2003) for a similar point. Brock et al. (1992) and Gencay (1998) verify the failure of the GARCH and GARCH-in-mean models to help improve conditional mean forecast.

neural networks combines many ‘basic’ nonlinear functions via a multilayer structure. Normally, there is one intermediate, or hidden, layer between the inputs and output. The intuition is that the explanatory variables simultaneously activate the units in the intermediate layer through some function Ψ and, subsequently, output is produced through some function Φ from the units in the intermediate layer. The following equations summarize this approach:

$$h_{i,t} = \Psi(\gamma_{i0} + \sum_{j=1}^{m} \gamma_{ij} X_{j,t}) \qquad i = 1,..., q \tag{1}$$

$$Y_t = \Phi(\beta_0 + \sum_{i=1}^{q} \beta_i h_{i,t}), \tag{2}$$

or more compactly,

$$Y_t = \Phi\left(\beta_0 + \sum_{i=1}^{q} \beta_i \Psi(\gamma_{i0} + \sum_{j=1}^{m} \gamma_{ij} X_{j,t}) \right), \tag{3}$$

where $X_{j,t}$ is the input or an independent variable, $h_{i,t}$ is the node or hidden unit in the intermediate or hidden layer, and Y_t is the output or dependent variable. In this study the independent variable $X_{j,t}$ coincides with the lagged dependent variable Y_{t-j}. The functions Ψ and Φ can be arbitrarily chosen and still approximate a large class of functions given sufficiently large numbers of units in the intermediate layer. Note that the correct number of lags needed is typically unknown, and in some instances lagged dependent variables may not be sufficient to capture the behavior of the time series.

As in Campbell, Lo, MacKinlay (1997) and Gencay (1998, 1999), we use single layer feedforward neural networks in this study, which is the most basic but perhaps most widely-used neural network in economic and financial applications. In this case the input variables are connected to multiple nodes (or hidden units), and at each node they are weighted (differently) and transformed by the same activation function Ψ. The output of each node is then weighted again by β_i and summed and transformed by a second activation function Φ. We chose the logistic function for the function Ψ and the identity function for the function Φ, which is also common practice in the literature (e.g., Campbell, Lo, MacKinlay, 1997; Gencay, 1998, 1999). It is well established theoretically that the single layer feedforward network, through the composition of a network of relatively simple functions, can approximate any complex nonlinear function to an arbitrary degree of accuracy with a suitable number of nodes or hidden units. Coefficients for the NN(d, q) model are estimated using nonlinear least squares via the Newton-Raphson algorithm. The final equation estimated is:

$$E(Y_t \mid I_{t-1}) = \beta_0 + \sum_{j=1}^{d} \beta_j Y_{t-j} + \sum_{i=1}^{q} \delta_i G(\gamma_{0i} + \sum_{j=1}^{d} \gamma_{ji} Y_{t-j}), \tag{4}$$

where $G(z) = (1 + e^{-z})^{-1}$ and is a function of Ψ, I_{t-1} is the information set available at $t-1$, and Y_t is the dependent variable (i.e., stock returns in this study).

2.2 The Functional Coefficient Model

The functional coefficient model, first proposed by Cai et al. (2000), is a new nonlinear time series model with state-dependent coefficients, which includes threshold autoregression models, smooth transition autoregression, and many other regime switching models as special cases. The basic model can be expressed as follows:

$$E(Y_t \mid I_{t-1}) = \alpha_0(U_t) + \sum_{j=1}^{d} \alpha_j(U_t) Y_{t-j} \tag{5}$$

where $\{(Y_t, U_t)'\}$ is a stationary process, and Y_t and U_t are scalar variables. It is important to choose an appropriate smoothing variable U_t. U_t may be chosen as a function of explanatory variable vector Y_{t-j} or as a function of other variables. In our forecasts of stock market index returns using past returns, U_t should be a certain combination of the lagged independent variables. There are several specific ways to choose U_t. Here we chose U_t as the difference between the log index price at time $t-1$ (p_{t-1}), and the moving average of the most recent periods L of the log prices at time $t-1$, or:

$$U_t = p_{t-1} - L^{-1}\sum_{j=1}^{L} p_{t-j} \tag{6}$$

We chose $L = 12$ to capture one-year moving average. Traders often use U_t as a buy or sell signal based on its sign, which reveals information on changes in direction. Thus, the FC model is well suited to forecasting the direction of stock price changes.

Similar to Cai et al. (2000), we estimate the term $\{a_j(U_t)\}$ nonparametrically using a local linear estimator. We approximate $a_j(U_t)$ locally (when U_j is close to u) by $a_j(U_t) = a_j + b_j(U_t - u)$. The local linear estimator at point u is $\hat{a}_j(u) = \hat{a}_j$, and $\{(\hat{a}_j, \hat{b}_j)\}$ are chosen by minimizing the sum of locally weighted squares defined as:

$$\sum_{t=1}^{N}[Y_t - a_j - b_j(U_t - u)]^2 K_h(U_t - u), \tag{7}$$

where $K_h(.)$ is the kernel function used as weights for points that are included to estimate $\{(\hat{a}_j, \hat{b}_j)\}$. We use the normal distribution as the kernel function, and h is the smoothing parameter or the bandwidth of the window of the kernel function, which is determined by the modified leave-one-out least square cross-validation method proposed in Cai et al. (2000). As pointed out in Harvey (2001), the choice of the kernel function would have little effect on nonparametric regression while h is the most important factor to be considered.

2.3 The Nonparametric Regression Model

It is well known that nonlinearities in the conditional mean may be very complicated and cannot be expressed explicitly. Hence, it may be desirable to use the nonparametric regression to estimate the model without specifying the forms of functions. Similar to Harvey (2001), we use the well- known kernel regression technique (with some improvements on bandwidth selection to maximize the forecasting power). In general, a nonparametric regression model can be generally expressed as:

$$E(Y_t \mid I_{t-1}) = g(Y_{t-1}, Y_{t-2}, ..., Y_{t-j}) . \tag{8}$$

As mentioned above with respect to the nonparametric estimator of $a_j(U_t)$ in the FC model, $g(.)$ can be estimated by local linear regression. At each point $y_t = \{ y_{t-1}, y_{t-2}, \ldots, y_{t-j} \}$, we can approximate $g(.)$ locally by a linear function $g(Y) = a + (Y - y)'b$. We can also approximate $g(y)$ locally simply by a constant function $g(Y) = a$ (i.e., the local constant estimator), which is the approach taken here. Compared to other estimators, it has also drawn the most theoretical attention and thus has clear theoretical properties for estimation and inference of nonparametric models. The local constant estimator at point y is given by $g(y) = \hat{a}$, where $\hat{a}$ minimizes the sum of local weighted squares:

$$\sum_{t=1}^{N} [Y_t - a]^2 \prod_{s=1}^{j} K_{hs} \ (Y_{t-s} - y_{t-s}) , \tag{9}$$

where $\prod_{s=1}^{j} K_{hs} \ (Y_{t-s} - y_{t-s})$ is the product kernel, $K_{hs}(.)$ is the univariate kernel function, and $h = (h_1, ... h_j)$ is chosen by the leave-one-out cross-validation procedure. As already noted, the smoothing parameter h is the most important parameter in nonparametric estimation. An inappropriately chosen h will give poor in-sample and out-of-sample prediction. Traditional nonparametric forecasting uses the h that minimizes the in-sample sum square errors to forecast the next-period value based on previous in-sample data. However, while this h is optimal for all in-sample data, it may not be the best h for out-of-sample forecasting. Consequently, we use a modified method to select the smoothing parameter.[9]

Our modified approach consists of finding the best h for out-of-sample forecasting and making forecasts based on this h^*.[10] In this procedure, we have two parameters to establish: (1) the out-of-sample evaluation length k is set equal to 20 ($\hat{x}_{81}$ to $\hat{x}_{100}$) in the example, and (2) the regression length m is set equal to 80 in the example. Hence, we denote the model as NP(k, m), where the parameters (k, m) are important to the forecasting performance of this modified nonparametric regression model. Stock market returns in different countries might have different time series properties. For a more (less) volatile time series, a shorter (longer) evaluation length (m) may be better and vise versa. We thus chose several different combinations of parameters (k, m) to search for the best forecasting performance. Nevertheless, although different combinations could affect the estimation results for a particular time series, it appears that its impact is not very substantial qualitatively for the cases considered in this study.

3. Empirical Results

3.1 Data Descriptions

Monthly return data for this study cover a 30-year period January 1975 to December 2004 for thirteen major international stock markets (except for Canada, for which the data start from January 1977): Australia (AU), Belgium (BE), Canada (CA), France (FR), Germany (GE),

---

[9] We thank Qi Li for making the suggestion.

[10] For example, suppose that we have data points of x_1 to x_{100} and that we want to forecast x_{101}. The traditional approach is to find the best h to minimize the 100 data points' in-sample sum of squared errors (based on x_1 to x_{100}) and then use this h^* and these data points (i.e., x_1 to x_{100}) to forecast x_{101}. We propose the following modified nonparametric forecasting methodology. We use h^* and data points of x_1 to x_{80} to forecast x_{81}, data points of x_2 to x_{81} are used to forecast x_{82}, …, data points of x_{20} to x_{99} are used to forecast x_{100}. We find the h^* that minimizes the sum of squared errors of out-of-sample forecast of points x_{81} to x_{100} and use this h^* and data points x_{21} to x_{100} to make our final forecast of x_{101}.

Hong Kong (HK), Italy (IT), Japan (JP), Netherlands (NE), Singapore (SI), Switzerland (SW), the United Kingdom (UK), and the United States (US). We exclude a few smaller markets considered by Patro and Wu (2004) because their growth and value stock return data are available for a very short sample period. We obtained the data of the market, growth, and value indexes for each country.[11] Value (growth) portfolios consist of firms of which the book-to-market ratio is among the highest (lowest) 30 percentile in a given market. We use returns denoted in both the local currency and the U.S. dollar and find qualitatively the same results. For brevity, we mainly focus on the results based on returns in denoted in the local currency.

We focus on monthly return data because international value and growth return data are available to us only at the monthly frequency. It is important to note that monthly data are actually more appropriate for the purpose of this paper than daily or weekly data. This is because the higher-frequency data are more vulnerable to market microstructure problems, e.g., non-synchronous trading in stocks, which can generate "artificial" intertemporal dependences in stock returns (Lo and MacKinlay, 1988; and Hsieh, 1991). Therefore, monthly data should provide cleaner evidence of return predictability than daily or weekly data.

3.2 The Results on Stock Market Indexes

We use the rolling technique to make out-of-sample forecasts. The rolling technique enables us to estimate the parameters of all the models using only a fixed-length window of past data rather than all previously available data. For example, suppose that there are N observations in total, where $N = R + P$, and P is the number of out-of-sample forecasts. In the basic rolling technique, we use the first R observations to forecast the return for period $R+1$, which we can denote P_1. Then we use observations from period 2 to period $R+1$ to forecast the return for

[11] We thank Ken French at Dartmouth College for making the data available through his website. The international data are originally from Morgan Stanley Capital International (MSCI) and the U.S. data are from CRSP (the Center

period $R+2$, or P_2 and so on. Swanson and White (1997) suggest that the rolling technique is plausible because it further allows for the (potentially nonlinear) relation between the current and past returns to evolve across time.

One needs to use a large number of observations to estimate nonlinear models, especially nonparametric nonlinear models. Therefore, we would like to make sure that the in-sample size (R) is relatively adequate for even parsimonious nonlinear models with only one independent variable (i.e., one lagged dependent variable in this study) (particularly the three nonparametric models: NN, FC, NP). On the other hand, the out-of-sample size (P) should also be adequate to detect the difference in forecasting performance. Hence, we consider $R:P=2:1$ as the ratio to have a good balance for the two considerations, which yields the in-sample size of 231 observations (with the exception of 215 observations for Canada) and the out-of-sample size of 115 observations (with the exception of 107 observations for Canada). According to Ashley (2003), the forecast evaluation tests can be most appropriately evaluated at the 10% significance level, given slightly more than 100 observations for out-of-sample forecast. Nevertheless, for robustness, we also experiment with alternative ratios of $R:P=3:1$ and $R:P=4:1$.

We use four forecasting evaluation criteria: (1) the mean squared forecast error (MSFE), (2) the mean absolute forecast error (MAFE), (3) the mean forecast trading return (MFTR) defined as $MFTR \equiv P^{-1}\sum_{t=R}^{n-1} sign(\hat{Y}_{t+1})Y_{t+1}$, where $sign(.)$ denotes $sign(\hat{Y}_{t+1})$ = 1 if $\hat{Y}_{t+1} \geq 0$ and $sign(\hat{Y}_{t+1})$ = -1 if $\hat{Y}_{t+1} < 0$, and (4) the mean correct forecast direction (MCFD) defined as $MCFD \equiv P^{-1}\sum_{t=R}^{n-1} 1\left[sign(\hat{Y}_{t+1})sign(Y_{t+1}) > 0\right]$, with $sign(.)$ defined as above. Because stock returns are volatile, forecast errors can be quite large from period to period. Therefore, the statistical

for Research in Security Prices) database. Patro and Wu (2004) find essentially the same results using the CRSP index and the MSCI index for the U.S.

accuracy of forecasts (as measured by MSFE and MSAE) does not necessarily imply economic accuracy in the sense of maximizing investor profits. For example, arguably, an investor is most interested in correctly forecasting changes in stock price movements; however, it is quite possible that wrong forecasts of price changes could have smaller MSFEs than correct forecasts of price changes. Granger (1992) emphasizes that, in this case, it is also desirable to compute economic measures of forecast accuracy, e.g., MFTR and MCFD.[12] Many other authors (e.g., Leitch and Tanner, 1991; Hong and Lee, 2003) have made similar points in the context of forecasting asset prices. In this regard, both MFTR and MCFD can be particularly informative to profit-maximizing investors. To summarize, the use of multiple criteria in this study provides comprehensive perspectives on the predictability of stock returns.

Tables 1 and 2 report the out-of-sample forecast results for thirteen international stock market indexes. Note that all these results are based on the use of one-period lagged returns only (i.e., d = 1) because the in-sample size of just over 200 observations only allows for one independent variable (i.e., one lagged returns in this study) for several nonparametric models. We also provide bootstrapped p-values for the forecast evaluation test of whether the difference between a forecasting model and the benchmark model is statistically significant.

Table 1 reports the results using statistical evaluation criteria MSFE and MAFE, which are in levels for the benchmark model, and in ratio relative to that of the benchmark model for the other forecasting models. The vast majority of the MSFE and MAFE ratios are greater than

[12] Closely following Fama (1991) and Gencay (1998, 1999), we do not explicitly allow for transaction costs in the evaluation of trading rule performance of various models. Although there are surely positive information and trading costs, and assuming zero information and trading costs is surely false, "… its advantage, however, is that it is a clean benchmark that allows me to sidestep the messy problem of deciding what are reasonable information and trading costs." (Fama, 1991, p. 1575). Different market participants could even have different transaction costs. According to Fama (1991), the researcher instead should focus on the more interesting task of laying out the evidence on the adjustment of prices to various kinds of information (e.g., past stock returns in this study). Here we also adopt this position, as our main interest lies in predictability rather than market efficiency. More importantly, however, note that predictability in most indexes, as measured by the direction of changes, should not be affected by this consideration of transaction costs.

one, and the associated *p*-values for MSFE and MAFE are almost always higher than the conventional significance level. The few exceptions are the AR model for Canada and the NN model for Singapore in terms of MSFE (both significant at the 10% level) and the NN model for Singapore in terms of MAFE (significant at the 5% level). Therefore, consistent with previous studies (e.g., Hsieh, 1991), the forecasting models cannot outperform the benchmark model in terms of statistical criteria. Also, the relative usefulness of the NN model compared to other nonparametric models for forecasting Singapore market is somewhat consistent with Gencay (1998). Nevertheless, although Gencay (1998) documents some evidence for significant nonlinear forecasting performance in terms of MSFE for daily Dow Jones Industrial Average indexes, we do not find such evidence for monthly CRSP value-weighted index and our finding is consistent with Hsieh (1991). This difference could be attributable to the pronounced nonsynchronous trading problem of daily price index, difference in the broad-basedness of the indexes, and difference in the sample periods, among others.

Table 2 presents the results based on economic criteria. We find that the market index in Singapore appears to be predictable. For example, The NN and NP models outperform the benchmark model in terms of MFTR at the 5% and 10% significance levels, respectively. Moreover, The NN and NP models yield substantially higher average trading returns (1.81% and 1.29% per month for NN and NP, respectively, during the out-of-sample period) than the benchmark model (0.47%). However, we do not uncover significant predictability for the other countries.

Table 2 also shows that, in terms of MCFD, in additional to very strong predictability of the direction of stock price changes for Singapore (i.e., 63% based on the NN model, which is significant at the 1% level, compared to 50% based on the benchmark), there is some (albeit marginal) evidence for predictability of the direction of stock price changes for Belgium (based on the FC model) and Japan (based on the NN model). Although the evidence is only marginally significant at the 10% level, the percentage of correct prediction of the price changes direction is indeed noticeably higher than that in the benchmark model, particularly for Japan (57% in the NN model versus 51% in the benchmark model).

To summarize, except for Belgium, Canada, Japan, and Singapore, major international stock markets appear to follow a random walk. We might attribute the predictability of Belgium, Canada, and Singapore to their relatively small market size. However, this explanation does not apply to Japan because it is the second-largest stock market.

The predictability for stock market indexes is often detected by the NN model. This result is consistent with widely perceived usefulness of artificial neural network in uncovering the nonlinearity-in-mean (e.g., Lee, White, and Granger, 1993). It is also noteworthy that none of the thirteen countries (except Italy) exhibits in-sample (linear) predictability in Patro and Wu (2004), who use the variance ratio test (see their Table 4). Therefore, our results appear to suggest that allowing for nonlinear models and multiple evaluation criteria might increase the power for uncovering predictability of international stock market indexes, although the evidence is not extremely strong.

3.3 The Results on Value Style Indexes

Tables 3 and 4 report the results for the value stock portfolios. We find significant predictability for value portfolios in three countries (Canada, Hong Kong, and the U.S.) in terms of both MSFE and MAFE (Table 3). Specifically, although the evidence for Canada is significant only at the 10% level in terms of either criterion, the forecast improvements can be substantial (with the MSFE ration of 0.91 based on the NP model). The predictability evidence for Hong Kong is significant at the 5% level in terms of MSFE but only at the 10% level in terms of MAFE. Somewhat surprisingly, such predictability is captured by the simple AR(1) model but not the more elaborate nonlinear models. Nevertheless, the forecast improvements do not appear to be impressive (with the ratios of 0.99). Interestingly, the most statistically significant evidence exists for the U.S. value stock portfolio. The evidence is significant at the 5% level in terms of both criteria, and there is a noticeable forecast improvement (with the MSFE ratio of 0.94 and the MAFE ratio of 0.96, both based on the NN model).

By using economic criteria, we find the confirming evidence for predictability in Hong Kong and U.S. value stock portfolios (Table 4). Specifically, the MCFD criterion shows that the FC model can somewhat improve the percentage of correct prediction of price changes directions for the Hong Kong value stock portfolio over the benchmark model (54% versus 50%), which is significant at the 10% level. The trading return for the U.S. (1.30% per month) based on the NN model is also significantly higher than the benchmark (1.14% per month) at the 10% significance level. Based on the same NN model, there is also some evidence (significant at the 10% level) for improved prediction of price changes directions for the U.S. value stock portfolio over the benchmark model.

Table 4 also shows that using economic criteria allows us to detect significant predictability of value stock portfolios in two more countries—Belgium and Singapore. The evidence for Belgium (based on the NP model) is significant at the 10% level in terms of both MFTR and MCFD. It is interesting to note that while we find that the general stock market for Singapore is predictable in terms of statistical criteria, there is no such evidence for its value stock portfolios. Nevertheless, the predictability evidence for the country shows up in terms of MFTR, which demonstrates the advantage of using both economic and statistical criteria. While the evidence is only significant at the 10% level, the associated return (2.21% per month based on the NN model) clearly dominates that (0.83% per month) of the benchmark model.

To summarize, based on the four evaluation criteria, there is evidence against the random walk hypothesis for value stock portfolios in five of thirteen international stock markets, namely, Belgium, Canada, Hong Kong, Singapore and the U.S. It is interesting to note that these five countries include three of the four countries of which the general stock markets are found to be unpredictable.[13] Therefore, in general, there is no clear evidence that value stock portfolios are more predictable than the market portfolio.

3.4 The Results on Growth Style Indexes

Table 5 shows that, for growth stock portfolios, there is no evidence that any forecasting model can outperform the benchmark model in terms of either MSFE or MAFE, with the exception of Belgium, for which both the AR and PN models outperform the benchmark model in terms of MAFE at the 5% level. However, Table 6 shows that we obtain quite different results if using economic criteria. In particular, we find evidence for predictability in growth stock portfolios in nine countries, including Belgium. This result clearly demonstrates the importance of considering economic criteria in the forecast evaluation, as Hong and Lee (2003) have emphasized in their exchange rate forecasting exercises. This result also confirms the argument of Clements and Smith (2001) that the forecast evaluation based on traditional statistical measures may fail to detect the superiority of the nonlinear models.

Specifically, there is predictability evidence in terms of MFTR in Table 6—all of which are significant at the 10% level—for Canada, Japan, Netherlands, Singapore, Switzerland, and the U.K. In particular, trading returns based on appropriate nonlinear models are far better than those of the benchmark model for Japan (0.68% versus -0.26%), Netherlands (0.92% versus 0.56), and Singapore (1.11% versus 0.28%). For MCFD, we have the confirming evidence that the percentage of correct prediction of price changes direction can be improved over the benchmark model for Canada, Hong Kong, Switzerland, and the U.K., which is significant at the

[13] The only exception is Japan, for which there is also some predictability evidence in terms of MCFD significant at the 15% level (with the p-value of 0.12 based on the PN model).

10% level. In addition, we find new predictability evidence for France at the 5% level.

To summarize, the forecasting models considered in this paper outperform the benchmark model for nine out of the thirteen markets, including Belgium, Canada, France, Hong Kong, Japan, Netherlands, Singapore, Switzerland, and the U.K. These nine countries include all the four countries of which the general stock markets show some predictability and four of the five countries of which the value stock portfolios show some predictability. Therefore, our results suggest that growth stock portfolios appear to be more predictable than both value stock portfolios and the general stock markets.

It is somewhat puzzling that growth stocks are not predictable in U.S. but in majority of the other developed countries. One possible explanation is that the forecasting models considered here have potential limitations, for example, they do not fully exploit the information available at the time of forecast. In particular, Guo and Savickas (2007) use theoretically motivated financial variables, i.e., stock market volatility and average idiosyncratic volatility, to forecast stock returns and find that the in-sample R-Squared is substantially higher for growth stocks than value stocks. Therefore, our international evidence confirms Guo and Savickas (2007)' finding that growth stocks are more predictable than value stocks. It is also interesting to note that, unlike the case for stock market indexes, the predictability for style stock indexes is no longer dominantly picked up by the NN model. Again, this result demonstrates the importance of considering several models rather than a single model (e.g., the neural network).

3.5 Robustness Check

We conduct the robustness check on the results based on alternative ratios of $R:P=3:1$ and $R:P=4:1$. For brevity, we only briefly summarize the main findings but detailed results are available on request. The ratio of $R:P=3:1$ yields the in-sample size of 260 observations and the out-of-sample size of 86 observations. The results based on this ratio are qualitatively the same as those reported above.

We further conduct the analysis based on the ratio of $R:P=4:1$, which yields the in-sample size of 277 observations and the out-of-sample size of 69 observations. Note that compared to the case of $R:P=2:1$ (115 observations or about 9.5 years), there is a substantial reduction in the out-of-sample size (69 observations or about 5.5 years). For a much shorter out-of-sample window, we find significant predictability in many countries. For example, in terms of at least one of the four evaluation criteria, stock market indexes are found to be predictable at the 5% significant level for two countries (Hong Kong and Japan) and at the 10% level for six countries (Australia, Belgium, Canada, Singapore, Switzerland, and the U.K.). Clearly, this evidence for stock market index predictability is much stronger than the results reported above and in previous studies (e.g., Patro and Wu, 2004). Similarly, growth stock indexes are found to be predictable at the 5% level for four countries (Australia, Canada, France, and Hong Kong) and at the 10% level for six countries (Belgium, Germany, Japan, Netherlands, Singapore, and the U.K.). In contrast, value stock indexes are found to be predictable at the 5% level for one country (Switzerland) and at the 10% level for two countries (Hong Kong and the U.S.). To summarize, although we find some interesting variations in the results based on the ratio of $R:P=4:1$, the main finding that growth stocks are predictable for most international stock markets remains qualitatively unchanged.

Patro and Wu (2004) also conduct supplementary analysis based on the return data denoted in the U.S. dollar. Obviously, using stock returns in the U.S. dollar compounds the effect of exchange rate changes. While Patro and Wu find somewhat stronger predictability for market indexes denoted in the U.S. dollar, their main finding on (un)predictability of monthly stock market indexes is unaffected. We reach a similar conclusion in our analysis; for brevity, these results are not reported here but are available on request.[14]

To address potential data snooping biases, we conduct White's (2000) test for out-of-sample multiple model comparisons. For brevity, we briefly summarize the main finding here but details are available on request. The majority of significant test statistics based on a single model remains qualitatively the same. This statement is particularly true, given the concerns of potentially low power of the test (e.g., Hong and Lee, 2003) and relatively small sample sizes of this study (which together would justify the use of somewhat higher significance levels, such as 15%). At any rate, the main finding of more predictability in growth stocks than in value stocks and the market is qualitatively unchanged.

Lastly, to address the concern whether the U.S. results are robust in different sample periods (e.g., Patro and Wu, 2004), we conduct further analysis using monthly U.S. data over the period July 1926 to December 2004.[15] In particular, we investigate two subsamples: July 1926 to June 1963 (with a total of 430 observations) and July 1963 to December 2004 (with a total of 484 observations). Table 7 shows that, based on the ratio $R:P=2:1$, the stock market index and growth stock index are unpredictable by any model during the first period. By contrast, the NN model detects predictability in the value stock index for each of the four criteria either at the 10% or 5% levels. However, Table 8 shows that this finding is somewhat sensitive to the alternative ratio of $R:P=4:1$, for which we find all three price indices to be predictable at the 10% level.

For the second subsample, the results (for brevity, not reported here) are qualitatively the same as those based on the data for the period January 1975 to December 2004. That is, regardless of the ratio of $R:P=2:1$ or $R:P=4:1$, the value stock index is predictable at the 10% level, especially when evaluated with economic criteria. However, we find negligible

[14] We also consider two-step and three-step ahead forecast using the US data. As might be expected, the forecast results for two- and three-step-ahead forecast performance is generally much weaker than the one-step-ahead forecast reported above.

predictability for the stock market index and growth stock index. To summarize, the additional results for the second subsample suggest that our main finding based on a relatively short sample period January 1975 to December 2004 appears to be quite reliable.

4. Conclusions

Using a model selection approach, this study investigates in the out-of-sample forecasting context the martingale behavior of growth and value style indexes as well as general stock market indexes for thirteen major international stock markets. In addition to the linear model, we also employ several popular nonparametric nonlinear models to capture potential nonlinearity-in-mean in stock returns. We find that growth stock portfolios appear to be more predictable than value stock portfolios as well as the general stock market indexes. By contrast, there is no clear evidence that value stock portfolios are more predictable than the general stock markets. To the best of our knowledge, the documented short-horizon predictability pattern based on international data is novel in the random walk tests literature. Our results also further provide out-of-sample and international evidence, in line with Guo and Savickas (2007).

Our results shed light on the recent debate about stock return predictability. Goyal and Welch (2006) reexamine the existent empirical evidence using the data updated to 2004 and find little support for the out-of-sample predictability. However, Cochrane (2006) argues that, because the dividend yield is quite volatile and the dividend growth is unpredictable, the dividend yield must forecast stock returns, especially at long horizons. In this paper, we address this issue using different assets (growth style indexes), different forecasting variables (past returns), different forecasting models (nonlinear models), and alternative forecasting evaluation criteria (economic criteria). Our analysis suggests that (growth) stock returns might be predictable.

[15] Campbell and Vuolteenaho (2004) also find that CAPM explains the value premium in the pre-1963 sample but

Also consistent with early authors (e.g., Leitch and Tanner, 1991; Hong and Lee, 2003), we emphasize the importance of using economic criteria, in addition to commonly used statistical criteria in the forecast evaluation. Such consideration appears to be crucial to the main finding of this study. In particular, while statistical criteria fail to reject the martingale hypothesis for all the growth stock price series except one country, economic criteria suggest predictability of the direction of price changes as well as trading returns for nine countries.

Our finding that growth stocks are more predictable than value stocks and the aggregate market might be consistent with behavioral models. For example, Barberis and Shleifer's (2003) style-level positive feedback model implies that predictability of style indexes can be as pronounced or even more pronounced than the market and Teo and Woo (2004) provide further supportive evidence for their model. Also, Lakonishok et al. (1994) suggest that growth and value stocks are predictable because of the correction for the mispricing: Initially, growth stocks tend to be overpriced and the value stocks tend to be underpriced. However, our results might also be consistent with rational pricing. In particular, growth stocks are more predictable than value stocks possibly because, as shown by Campbell and Vuolteenaho (2004), the former is more sensitive to the discount rate shock, which has only a temporary effect on stock prices. In this paper, we do not try to differentiate the two alternative explanations because our main focus is stock return predictability. Nevertheless, it will be interesting to address this issue in future research.

References

Abraham, A., Seyyed, F. J., Alsakran, S.A., 2002. Testing the random walk behavior and efficiency of the Gulf stock markets. Financial Review 37, 469-480.

not the post-1963 sample.

Al-Khazali, O.M., Ding, D.K., and Pyun, C.S., 2007. A new variance ratio test of random walk in emerging markets: A revisit. Financial Review 42, forthcoming.

Ashley, R., 2003. Statistically significant forecasting improvements: How much out-of-sample data is likely necessary? International Journal of Forecasting 19, 229–239.

Ayadi, O.F., and Pyun, C.S., 1994. The application of the variance ratio test to the Korean securities market. Journal of Banking and Finance 18, 643-658.

Barberis, N., and Shleifer, A., 2003. Style investing. Journal of Financial Economics 68, 161-199.

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

Malkiel, B.G., 2005. Reflections on the Efficient Market Hypothesis: 30 Years Later Financial Review 40,1-9.

Cai, Z., Fan, J., and Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of American Statistical Association 95, 941-956.

Campbell, J. and J. Cochrane, 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205-251.

Campbell, J., Lo, A., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, New Jersey.

Campbell, J., and Vuolteenaho T., 2004. Bad Beta, Good Beta, American Economic Review 94, 1249-1275.

Chaudhuri, K., and Wu, Y., 2003. Random walk versus breaking trend in stock prices: Evidence from emerging markets. Journal of Banking and Finance 27, 575-592.

Clements, M.P., and Smith, J., 2001. Evaluating forecasts from SETAR models of exchange rates. Journal of International Money and Finance 20, 133-148.

Chordia, T., Roll, R., and Subrahmanyam, A., 2005. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics 76, 271-292.

Cochrane, J., 2006. A defense of return predictability. Unpublished Working Paper, University of Chicago.

Coggin, T. D., 1998. Long-term memory in equity style indexes. Journal of Portfolio Management 24 (2), 37-46.

Darrat, A., Zhong, M., 2000. On testing the random walk hypothesis: A model-comparison approach, Financial Review 35, 105-124.

Fama, E.F., 1965. The behavior of stock market prices. Journal of Business 38, 34-105.

Fama, E.F., 1991. Efficient capital markets: II. Journal of Finance 46, 1575-1617.

Gencay, R., 1998. The predictability of security returns with simple technical trading rules. Journal of Empirical Finance 5, 347-359.

Gencay, R., 1999. Linear, nonlinear and essential foreign exchange rate prediction with simple trading rules. Journal of International Economics 47, 91-107.

Goyal, A. and Welch, I., 2006. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies, forthcoming.

Granger, C.W.J., 1992. Forecasting stock market prices: Lessons for forecasters. International Journal of Forecasting 8, 3-13.

Guo, H. and Savickas, R., 2007. Aggregate idiosyncratic volatility in G7 countries. Review of Financial Studies, forthcoming.

Harvey, C.R., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573-637.

Hogan, S., Jarrow, R., Teo, M., and Warachka, M., 2004. Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. Journal of Financial Economics 73, 525-565.

Hong, Y.M., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. Journal of American Statistical Association 84, 1201-1220.

Hong, Y., Chung, J. 2003. Are the directions of stock price changes predictable? Statistical theory and evidence, Working Paper, Cornell University.

Hong, Y. M., and Lee, T. H., 2003. Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. Review of Economics and Statistics 85, 1048-1062.

Hsieh, D.A., 1991. Chaos and nonlinear dynamics: Application to financial markets. Journal of Finance 46, 1839-1877.

Lakonishok, J., Sheleifer, A., and Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541-1578.

Lee, T.H., White, H., and Granger, C.W.J. 1993. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. Journal of Econometrics 56, 269-290.

Leitch, G., and Tanner, E., 1991. Economic forecast evaluation: profits versus conventional error measures. American Economic Review 81, 580-590.

Lo, A.W., and Mackinlay, A.C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Review of Financial studies 1, 41-66.

Mcqueen, G., and Thorley, S., 1991. Are stock returns predictable? A test using Markov chains. Journal of Finance 46, 239-263.

Patro, D.K., and Wu, Y., 2004. Predictability of short-horizon returns in international equity markets. Journal of Empirical Finance 11, 553-584.

Swanson, N.R., and White, H., 1995. A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business Economics and Statistics 13, 265-275.

Swanson, N.R., and White, H., 1997. A model selection approach to real time macroeconomic forecasting using linear models and artificial neural networks. Review of Economics and Statistics 79, 540-550.

Teo, M., and Woo, S.J., 2004. Style effects in the cross-section of stock returns. Journal of Financial Economics 74, 367-398.

Urrutia, J.L., 1995. Tests of random walk and market efficiency for Latin American emerging equity markets. Journal of Financial Research 18, 299-309.

Urrutia, J.L., Vu, J., Gronewoller, P., Hoque, M., 2002. Nonlinearity and Low Deterministic Chaotic Behavior in Insurance Portfolio Stock Returns. Journal of Risk & Insurance 69, 537–554.

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

Wright, J.H., 2000. Alternative variance-ratio tests using ranks and signs, Journal of Business and Economic Statistics 18, 1-9.

Table 1 Forecast Evaluation Results for Stock Market Indexes – MSFE and MAFE

	AU	BE	CA	FR	GE	HK	IT	JP	NE	SI	SW	UK	US
							Panel A. MSFE						
Benchmark	10.71	23.16	20.83	34.11	43.88	65.01	48.47	24.83	32.13	50.93	23.99	16.34	22.33
AR	1.01	0.99	**0.98**	1.00	1.01	1.00	1.02	1.00	1.01	1.00	0.99	1.00	1.00
	(.88)	(.33)	**(.09)**	(.52)	(.73)	(.49)	(.91)	(.28)	(.66)	(.47)	(.29)	(.78)	(.77)
PN	1.02	1.01	1.02	1.00	1.17	1.01	1.06	1.04	1.08	1.05	0.99	1.04	1.02
	(.82)	(.61)	(.93)	(.55)	(.90)	(.75)	(.96)	(.88)	(.98)	(.87)	(.34)	(.94)	(.86)
NN	1.03	0.98	1.04	1.08	1.04	1.10	0.98	1.08	1.00	**0.96**	1.06	1.09	1.01
	(.80)	(.18)	(.93)	(.94)	(.71)	(.90)	(.29)	(.87)	(.45)	**(.08)**	(.75)	(.98)	(.63)
FC	1.04	1.15	1.10	1.04	1.20	1.09	1.04	1.04	1.64	1.83	2.24	1.08	1.41
	(.96)	(.95)	(.93)	(.77)	(.93)	(.72)	(.77)	(.87)	(.92)	(.90)	(.89)	(.80)	(.95)
NP	1.02	1.00	1.05	1.05	1.03	1.01	1.02	0.98	1.08	1.01	0.96	1.06	1.01
	(.86)	(.54)	(.96)	(.87)	(.80)	(.55)	(.93)	(.34)	(.89)	(.62)	(.12)	(.88)	(.82)
							Panel B. MAFE						
Benchmark	2.54	3.59	3.49	4.45	4.95	5.96	5.42	4.11	4.22	5.14	3.65	3.00	3.76
AR	1.00	0.98	1.00	1.00	1.00	1.00	1.02	1.00	1.01	1.02	1.00	1.01	1.00
	(.74)	(.15)	(.17)	(.35)	(.74)	(.70)	(.96)	(.21)	(.90)	(.97)	(.40)	(.94)	(.75)
PN	1.02	0.98	1.01	1.01	1.05	1.00	1.03	1.01	1.03	1.03	1.00	1.02	1.01
	(.94)	(.16)	(.85)	(.82)	(.88)	(.59)	(.95)	(.72)	(.97)	(.93)	(.52)	(.93)	(.69)
NN	1.01	0.99	1.03	1.05	1.02	1.07	1.02	1.01	1.01	**0.96**	1.01	1.05	1.00
	(.59)	(.40)	(.97)	(.94)	(.76)	(.99)	(.78)	(.60)	(.61)	**(.04)**	(.65)	(.97)	(.52)
FC	1.02	1.06	1.04	1.02	1.07	1.05	1.01	1.02	1.19	1.17	1.14	1.05	1.13
	(.92)	(.94)	(.85)	(.80)	(.93)	(.80)	(.62)	(.76)	(.97)	(.97)	(.85)	(.87)	(.97)
NP	1.02	1.00	1.03	1.02	1.02	1.00	1.01	0.98	1.02	1.02	0.98	1.01	1.00
	(.99)	(.46)	(.91)	(.82)	(.85)	(.56)	(.86)	(.23)	(.86)	(.73)	(.14)	(.67)	(.54)

Note: The monthly data cover the period January 1975 to December 2004. The actual values of MSFE and MAFE are reported for the benchmark model (i.e., the martingale). The MSFE and MAFE for other models are the ratios to the MSPE and MAFE of the benchmark model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 2 Forecast Evaluation Results for Stock Market Indexes – MFTR and MCFD

	AU	BE	CA	FR	GE	HK	IT	JP	NE	SI	SW	UK	US
							Panel A. MFTR						
Benchmark	0.98	1.10	1.01	1.06	0.86	0.86	1.18	0.20	0.93	0.47	0.89	0.70	0.97
AR	0.98	1.03	1.01	0.72	0.61	0.86	1.18	0.20	0.69	-0.07	0.72	0.70	0.97
	(.96)	(.71)	(.94)	(.88)	(.88)	(.96)	(.94)	(.94)	(.85)	(.95)	(.76)	(.94)	(.96)
PN	0.98	1.00	0.96	0.82	0.67	0.90	1.11	-0.10	0.69	0.13	0.89	0.70	0.97
	(.97)	(.73)	(.89)	(.71)	(.75)	(.24)	(.60)	(.74)	(.87)	(.68)	(.42)	(.94)	(.94)
NN	0.98	1.18	0.90	1.03	1.03	-0.07	1.29	0.35	1.14	**1.81**	1.13	0.46	0.92
	(.94)	(.41)	(.77)	(.51)	(.39)	(.95)	(.16)	(.43)	(.20)	**(.04)**	(.29)	(.86)	(.61)
FC	0.98	1.25	1.01	0.81	1.35	0.43	1.35	-0.14	1.20	0.58	0.76	0.89	0.47
	(.95)	(.35)	(.94)	(.67)	(.13)	(.83)	(.32)	(.69)	(.26)	(.46)	(.66)	(.28)	(.99)
NP	0.97	1.00	0.99	0.63	0.89	1.27	0.97	0.14	0.88	1.29	1.00	0.55	0.99
	(.73)	(.74)	(.73)	(.92)	(.44)	(.13)	(.75)	(.54)	(.62)	(.07)	(.24)	(.74)	(.35)
							Panel B. MCFD						
Benchmark	0.68	0.64	0.63	0.63	0.62	0.54	0.51	0.51	0.66	0.50	0.64	0.63	0.63
AR	0.68	0.66	0.63	0.62	0.60	0.54	0.51	0.51	0.64	0.46	0.63	0.63	0.63
	(.94)	(.15)	(.96)	(.64)	(.86)	(.96)	(.94)	(.97)	(.85)	(.96)	(.73)	(.96)	(.96)
PN	0.68	0.63	0.62	0.62	0.61	0.54	0.50	0.53	0.64	0.52	0.63	0.63	0.63
	(.96)	(.73)	(.64)	(.73)	(.57)	(.34)	(.63)	(.35)	(.86)	(.31)	(.64)	(.94)	(.96)
NN	0.68	0.64	0.62	0.59	0.60	0.51	0.52	**0.57**	0.66	**0.63**	0.66	0.56	0.63
	(.92)	(.43)	(.62)	(.80)	(.69)	(.75)	(.28)	**(.10)**	(.32)	**(.01)**	(.29)	(.99)	(.45)
FC	0.68	**0.67**	0.63	0.62	0.63	0.53	0.53	0.54	0.63	0.50	0.61	0.63	0.57
	(.96)	**(.10)**	(.96)	(.53)	(.29)	(.58)	(.22)	(.33)	(.76)	(.44)	(.88)	(.43)	(.99)
NP	0.67	0.63	0.62	0.60	0.61	0.57	0.52	0.53	0.66	0.55	0.64	0.63	0.63
	(.72)	(.75)	(.74)	(.93)	(.59)	(.15)	(.26)	(.33)	(.33)	(.11)	(.32)	(.73)	(.35)

Note: The monthly data cover the period January 1975 to December 2004. The benchmark model is the martingale model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 3 Forecast Evaluation Results for Value Stock Indexes – MSFE and MAFE

	AU	BE	CA	FR	GE	HK	IT	JP	NE	SI	SW	UK	US
							Panel A. MSFE						
Benchmark	11.98	31.20	37.63	55.62	59.15	130.83	74.33	52.24	65.34	101.32	61.49	28.66	17.74
AR	1.00	0.99	1.01	1.01	1.02	**0.98**	1.01	1.00	0.99	0.99	0.96	1.00	0.99
	(.47)	(.42)	(.73)	(.55)	(.94)	**(.02)**	(.68)	(.65)	(.40)	(.35)	(.15)	(.59)	(.22)
PN	1.03	1.04	0.97	1.06	1.13	1.00	1.05	1.30	1.03	1.65	1.01	1.03	1.02
	(.97)	(.77)	(.20)	(.79)	(.91)	(.51)	(.92)	(.86)	(.71)	(.79)	(.61)	(.83)	(.89)
NN	0.97	0.97	1.02	1.14	1.12	0.99	1.05	1.01	1.12	0.95	1.05	1.12	**0.94**
	(.15)	(.24)	(.80)	(.93)	(.91)	(.36)	(.83)	(.73)	(.93)	(.20)	(.72)	(.99)	**(.03)**
FC	1.07	1.07	1.17	1.65	1.16	3.04	1.13	5.45	1.24	1.99	2.20	1.67	1.04
	(.97)	(.76)	(.97)	(.89)	(.92)	(.89)	(.83)	(.92)	(.94)	(.89)	(.96)	(.97)	(.73)
NP	1.02	1.03	**0.91**	1.11	1.04	0.98	1.02	1.06	1.06	1.07	0.97	1.04	1.01
	(.76)	(.85)	**(.06)**	(.86)	(.89)	(.17)	(.88)	(.95)	(.79)	(.89)	(.18)	(.89)	(.82)
							Panel B. MAFE						
Benchmark	2.58	4.29	4.75	5.31	5.43	8.17	6.47	5.61	5.97	6.70	5.60	4.04	3.24
AR	1.00	1.00	1.01	1.02	1.01	**0.98**	1.01	1.00	0.99	1.01	0.99	1.00	1.00
	(.37)	(.52)	(.92)	(.88)	(.85)	**(.09)**	(.89)	(.31)	(.25)	(.60)	(.28)	(.51)	(.28)
PN	1.02	1.02	1.00	1.05	1.06	1.00	1.05	1.05	1.01	1.13	1.00	1.01	1.00
	(.99)	(.71)	(.37)	(.95)	(.94)	(.48)	(.96)	(.83)	(.60)	(.93)	(.57)	(.82)	(.60)
NN	0.99	1.00	1.02	1.09	1.06	1.02	1.05	0.99	1.07	1.02	1.01	1.04	**0.96**
	(.20)	(.44)	(.86)	(.98)	(.96)	(.80)	(.98)	(.36)	(.96)	(.72)	(.63)	(.92)	**(.04)**
FC	1.05	1.05	1.10	1.16	1.06	1.21	1.03	1.26	1.06	1.18	1.31	1.16	1.03
	(.99)	(.87)	(.99)	(.94)	(.88)	(.90)	(.76)	(.89)	(.90)	(.93)	(.99)	(.98)	(.83)
NP	1.01	1.01	**0.96**	1.06	1.03	0.99	1.03	1.02	1.02	1.05	0.98	1.02	1.00
	(.70)	(.79)	**(.06)**	(.92)	(.93)	(.32)	(.97)	(.89)	(.73)	(.96)	(.18)	(.87)	(.63)

Note: The monthly data cover the period January 1975 to December 2004. The actual values of MSFE and MAFE are reported for the benchmark model (i.e., the martingale). The MSFE and MAFE for other models are the ratios to the MSPE and MAFE of the benchmark model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 4 Forecast Evaluation Results for Value Stock Indexes – MFTR and MCFD

	AU	BE	CA	FR	GE	HK	IT	JP	NE	SI	SW	UK	US
							Panel A. MFTR						
Benchmark	1.26	1.82	0.97	1.31	1.68	0.97	1.10	1.23	1.72	0.83	1.01	0.81	1.14
AR	1.26	1.65	0.87	1.49	1.68	0.83	1.21	1.09	1.88	0.51	1.79	0.56	1.07
	(.96)	(.88)	(.89)	(.33)	(.96)	(.79)	(.35)	(.74)	(.43)	(.67)	(.17)	(.74)	(.72)
PN	1.26	1.82	0.49	0.84	1.66	0.55	1.10	1.04	1.91	0.26	1.48	0.56	1.07
	(.94)	(.95)	(.91)	(.81)	(.52)	(.75)	(.52)	(.67)	(.38)	(.79)	(.25)	(.73)	(.75)
NN	1.25	1.64	1.10	-0.01	1.09	-0.02	0.35	1.04	1.14	**2.21**	0.92	0.30	**1.30**
	(.74)	(.71)	(.33)	(.97)	(.84)	(.78)	(.96)	(.79)	(.76)	**(.06)**	(.55)	(.94)	**(.08)**
FC	1.21	1.72	0.05	0.64	0.71	0.79	1.40	0.01	2.01	-0.94	1.49	0.40	1.03
	(.74)	(.57)	(.97)	(.80)	(.96)	(.61)	(.32)	(.94)	(.28)	(.90)	(.28)	(.91)	(.65)
NP	1.20	**1.86**	1.48	1.07	1.98	1.37	0.94	0.10	1.73	0.46	0.90	0.75	1.14
	(.75)	**(.08)**	(.25)	(.66)	(.16)	(.27)	(.65)	(.91)	(.47)	(.81)	(.66)	(.62)	(.96)
							Panel B. MCFD						
Benchmark	0.72	0.62	0.62	0.64	0.65	0.50	0.51	0.57	0.67	0.51	0.61	0.60	0.62
AR	0.72	0.62	0.60	0.65	0.65	0.49	0.50	0.56	0.66	0.49	0.63	0.59	0.63
	(.96)	(.85)	(.87)	(.31)	(.96)	(.60)	(.65)	(.74)	(.59)	(.75)	(.25)	(.75)	(.74)
PN	0.72	0.63	0.58	0.62	0.65	0.50	0.50	0.58	0.67	0.47	0.63	0.59	0.63
	(.94)	(.96)	(.93)	(.79)	(.45)	(.44)	(.56)	(.12)	(.41)	(.86)	(.23)	(.75)	(.74)
NN	0.71	0.62	0.62	0.59	0.59	0.51	0.46	0.55	0.62	0.55	0.62	0.55	**0.64**
	(.73)	(.65)	(.44)	(.92)	(.95)	(.37)	(.86)	(.61)	(.91)	(.21)	(.37)	(.97)	**(.07)**
FC	0.71	0.60	0.51	0.61	0.58	**0.54**	0.53	0.50	0.68	0.46	0.62	0.58	0.60
	(.74)	(.85)	(.99)	(.77)	(.98)	**(.06)**	(.29)	(.92)	(.24)	(.84)	(.36)	(.76)	(.93)
NP	0.71	**0.64**	0.64	0.63	0.66	0.50	0.48	0.55	0.66	0.50	0.60	0.61	0.63
	(.61)	**(.08)**	(.16)	(.56)	(.22)	(.43)	(.81)	(.65)	(.62)	(.73)	(.57)	(.96)	(.17)

Note: The monthly data cover the period January 1975 to December 2004. The benchmark model is the martingale model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 5 Forecast Evaluation Results for Growth Stock Indexes – MSFE and MAFE

	AU	BE	CA	FR	GE	HK	IT	JP	NE	SI	SW	UK	US
							Panel A. MSFE						
Benchmark	12.73	26.71	42.38	36.05	54.53	41.37	50.83	27.87	24.61	52.29	18.83	15.44	25.18
AR	1.01	0.98	1.01	0.99	1.01	1.00	1.02	1.00	1.01	1.01	0.98	1.00	1.01
	(.93)	(.11)	(.81)	(.30)	(.74)	(.57)	(.96)	(.73)	(.93)	(.83)	(.17)	(.75)	(.79)
PN	1.02	0.99	1.04	0.98	1.08	1.01	1.02	1.02	1.01	1.02	0.99	1.06	1.02
	(.93)	(.24)	(.85)	(.20)	(.93)	(.76)	(.89)	(.81)	(.71)	(.84)	(.31)	(.99)	(.82)
NN	1.17	0.98	1.09	0.99	1.03	0.97	1.12	1.05	0.95	0.98	1.06	1.12	1.01
	(.98)	(.33)	(.92)	(.43)	(.70)	(.29)	(.99)	(.89)	(.11)	(.37)	(.84)	(.99)	(.62)
FC	1.07	1.08	2.19	1.25	1.18	1.01	1.03	1.76	1.15	1.29	1.81	2.21	2.37
	(.99)	(.83)	(.97)	(.76)	(.97)	(.54)	(.71)	(.91)	(.97)	(.97)	(.87)	(.94)	(.91)
NP	1.03	1.05	1.09	1.00	1.05	1.08	1.04	1.00	1.00	1.03	1.00	1.02	1.05
	(.78)	(.83)	(.93)	(.49)	(.83)	(.92)	(.88)	(.61)	(.70)	(.82)	(.56)	(.88)	(.88)
							Panel B. MAFE						
Benchmark	2.77	3.93	4.99	4.56	5.26	4.66	5.61	4.34	3.68	5.20	3.27	2.90	4.01
AR	1.01	**0.98**	0.99	0.99	1.00	1.00	1.02	1.00	1.01	1.01	0.99	1.01	1.01
	(.91)	**(.04)**	(.20)	(.16)	(.50)	(.82)	(.98)	(.81)	(.98)	(.94)	(.17)	(.93)	(.92)
PN	1.02	**0.98**	1.01	0.99	1.03	1.00	1.01	1.01	1.02	1.01	1.00	1.04	1.02
	(.98)	**(.03)**	(.81)	(.29)	(.89)	(.50)	(.86)	(.79)	(.87)	(.83)	(.36)	(.99)	(.95)
NN	1.10	0.99	1.02	1.01	1.02	1.00	1.06	1.02	0.98	1.03	1.01	1.06	1.01
	(.99)	(.38)	(.75)	(.65)	(.75)	(.53)	(.99)	(.80)	(.27)	(.89)	(.67)	(.99)	(.71)
FC	1.05	1.05	1.29	1.02	1.08	1.02	1.03	1.13	1.10	1.12	1.10	1.24	1.19
	(.99)	(.85)	(.99)	(.68)	(.95)	(.62)	(.88)	(.91)	(.99)	(.98)	(.87)	(.96)	(.91)
NP	1.04	1.01	1.03	1.00	1.02	1.04	1.00	0.99	1.02	1.03	1.01	1.01	1.02
	(.91)	(.78)	(.88)	(.47)	(.76)	(.94)	(.51)	(.27)	(.95)	(.89)	(.66)	(.87)	(.83)

Note: The monthly data cover the period January 1975 to December 2004. The actual values of MSFE and MAFE are reported for the benchmark model (i.e., the martingale). The MSFE and MAFE for other models are the ratios to the MSPE and MAFE of the benchmark model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 6 Forecast Evaluation Results for Growth Stock Indexes – MFTR and MCFD

	AU	BE	CA	FR	GE	HK	IT	JP	NE	SI	SW	UK	US
							Panel A. MFTR						
Benchmark	0.72	0.94	0.68	0.96	0.76	0.92	1.06	-0.26	0.56	0.28	0.90	0.64	0.95
AR	0.72	1.01	**0.70**	0.76	0.07	0.92	1.06	-0.23	0.56	0.28	0.85	0.64	0.95
	(.96)	(.11)	**(.08)**	(.75)	(.90)	(.96)	(.96)	(.41)	(.96)	(.96)	(.59)	(.97)	(.96)
PN	0.70	1.00	0.80	0.83	0.15	0.81	1.06	-0.13	0.53	-0.21	**1.02**	0.64	0.95
	(.74)	(.12)	(.23)	(.75)	(.87)	(.76)	(.94)	(.43)	(.74)	(.87)	**(.09)**	(.96)	(.94)
NN	0.47	0.99	0.50	0.78	1.25	1.27	0.32	0.22	**0.92**	**1.11**	0.87	0.54	0.91
	(.79)	(.47)	(.61)	(.78)	(.22)	(.22)	(.90)	(.25)	**(.08)**	**(.10)**	(.57)	(.93)	(.53)
FC	0.72	0.96	1.16	1.07	0.14	0.92	1.04	**0.65**	0.70	0.38	0.60	0.41	0.54
	(.96)	(.47)	(.25)	(.44)	(.91)	(.96)	(.50)	**(.10)**	(.30)	(.44)	(.81)	(.84)	(.89)
NP	0.56	0.72	0.40	1.31	1.18	0.64	0.83	**0.68**	0.53	0.26	0.81	**0.65**	0.81
	(.84)	(.86)	(.82)	(.15)	(.25)	(.84)	(.72)	**(.06)**	(.74)	(.52)	(.66)	**(.08)**	(.74)
							Panel B. MCFD						
Benchmark	0.57	0.63	0.56	0.63	0.60	0.59	0.52	0.51	0.62	0.54	0.62	0.60	0.61
AR	0.57	0.64	**0.58**	0.62	0.59	0.59	0.52	0.50	0.62	0.54	0.62	0.60	0.61
	(.91)	(.11)	**(.09)**	(.73)	(.65)	(.96)	(.94)	(.57)	(.96)	(.93)	(.61)	(.96)	(.94)
PN	0.56	0.63	0.57	0.62	0.59	0.59	0.52	0.47	0.61	0.51	**0.64**	0.60	0.61
	(.73)	(.19)	(.33)	(.65)	(.56)	(.34)	(.96)	(.74)	(.74)	(.81)	**(.09)**	(.92)	(.96)
NN	0.57	0.62	0.56	0.60	0.62	**0.62**	0.50	0.52	0.63	0.53	0.63	0.57	0.59
	(.48)	(.57)	(.59)	(.82)	(.31)	**(.09)**	(.68)	(.40)	(.20)	(.52)	(.42)	(.92)	(.70)
FC	0.57	0.62	0.61	0.68	0.57	0.59	0.50	0.57	0.60	0.54	0.61	0.59	0.56
	(.86)	(.58)	(.22)	(.12)	(.73)	(.94)	(.84)	(.21)	(.69)	(.47)	(.69)	(.58)	(.95)
NP	0.53	0.61	0.55	**0.67**	0.60	0.57	0.53	0.56	0.61	0.52	0.62	**0.62**	0.61
	(.86)	(.72)	(.74)	**(.05)**	(.39)	(.81)	(.33)	(.19)	(.74)	(.68)	(.63)	**(.08)**	(.42)

Note: The monthly data cover the period January 1975 to December 2004. The benchmark model is the martingale model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 7 Forecast Evaluation Results for the US: July 1926 to June 1963 and R:P = 2:1

(R,P) = (287,143)		MSFE		MAFE		MFTR		MCFD	
k	Model	MSFE	*P*	MAFE	*P*	MFTR	*P*	MCFD	*P*
				Panel A. The Market Index					
0	Benchmark	12.43		2.82		1.15		0.67	
1	AR	1.02	.93	1.00	.67	1.10	.86	0.66	.86
2	PN	1.03	.99	1.01	.92	1.14	.75	0.66	.74
3	NN	0.99	.31	1.00	.50	1.14	.71	0.66	.76
4	FC	1.03	.75	1.01	.75	1.05	.87	0.66	.86
5	NP	1.05	.87	1.02	.91	1.17	.33	0.67	.33
				Panel B. The Value Stock Index					
0	Benchmark	17.77		3.23		1.36		0.65	
1	AR	1.04	.92	1.02	.91	**1.39**	**.09**	**0.67**	**.07**
2	PN	1.05	.98	1.03	.99	**1.39**	**.08**	**0.67**	**.08**
3	NN	1.06	.89	1.03	.89	1.33	.57	0.65	.80
4	FC	1.06	.83	1.04	.91	1.18	.83	0.65	.77
5	NP	1.05	.91	1.04	.99	1.28	.74	0.64	.92
				Panel C. The Growth Stock Index					
0	Benchmark	14.19		3.03		1.11		0.65	
1	AR	1.01	.91	1.00	.72	1.11	.99	0.65	.96
2	PN	1.02	.93	1.01	.96	0.94	.90	0.64	.79
3	NN	1.04	.87	1.01	.75	0.90	.86	0.63	.71
4	FC	1.10	.89	1.03	.89	0.89	.83	0.62	.93
5	NP	1.01	.58	1.01	.73	1.20	.28	0.66	.16

Note: The actual values of MSFE and MAFE are reported for the benchmark model (i.e., the martingale). The MSFE and MAFE for other models are the ratios to the MSPE and MAFE of the benchmark model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.

Table 8 Forecast Evaluation Results for the US: July 1926 to June 1963 and R:P = 4:1

(R,P) = (344,86)		MSFE		MAFE		MFTR		MCFD	
k	Model	MSFE	*P*	MAFE	*P*	MFTR	*P*	MCFD	*P*
				Panel A. The Market Index					
0	Benchmark	13.76		2.97		0.80		0.66	
1	AR	0.99	.43	1.00	.55	**0.90**	**.08**	**0.67**	**.08**
2	PN	0.98	.22	1.01	.78	0.78	.76	0.65	.73
3	NN	1.02	.69	1.02	.81	0.68	.72	0.65	.75
4	FC	1.04	.87	1.02	.82	0.67	.81	0.64	.82
5	NP	1.05	.75	1.01	.74	0.84	.32	0.66	.31
				Panel B. The Value Stock Index					
0	Benchmark	18.00		3.23		1.00		0.64	
1	AR	0.97	.20	**0.95**	**.03**	1.01	.48	0.64	.25
2	PN	0.98	.24	0.99	.29	**1.01**	**.08**	**0.65**	**.08**
3	NN	0.99	.48	0.99	.37	1.19	.23	0.66	.16
4	FC	0.97	.34	0.98	.25	0.65	.83	0.61	.76
5	NP	0.99	.43	0.99	.42	0.93	.74	0.63	.73
				Panel C. The Growth Stock Index					
0	Benchmark	15.86		3.20		0.72		0.62	
1	AR	1.00	.46	1.00	.46	**0.81**	**.08**	**0.63**	**.09**
2	PN	0.99	.37	1.01	.77	0.71	.53	0.61	.68
3	NN	1.05	.83	1.01	.65	0.73	.48	0.58	.75
4	FC	1.16	.90	1.06	.92	0.41	.80	0.59	.73
5	NP	1.10	.83	1.02	.72	0.86	.29	0.63	.18

Note: The actual values of MSFE and MAFE are reported for the benchmark model (i.e., the martingale). The MSFE and MAFE for other models are the ratios to the MSPE and MAFE of the benchmark model. The numbers in the parentheses are bootstrapped p-values for testing the significant difference in the forecasting performance between each of the models and the benchmark. The statistics which are significant at the 10% level or lower are in bold in the table.