Emerging Markets Group

EMG Working Paper Series

WP-EMG-26-2009

'Out of Sample Predictability in International Equity Markets: A Model Selection Approach'

Xiaojing Su, Tao Wang and Jian Yang

November 2009

Emerging Markets Group
Cass Business School
City University
106 Bunhill Row
London
EC1Y 8TZ
UK

www.cass.city.ac.uk/emg/

Out-of-Sample Predictability in International Equity Markets: A Model Selection Approach

Xiaojing Su
Hong Kong Monetary Authority

Tao Wang
Queens College and the Graduate Center of CUNY

Jian Yang*
University of Colorado Denver

Forthcoming, *Financial Review*

Abstract

For thirteen major international stock markets, there is much evidence of out-of-sample predictability for growth stocks especially when evaluated with economic criteria, and to a noticeably lesser extent for general stock markets and value stocks. Our results shed light on the recent debate about stock return predictability, using different assets (growth style indexes), forecasting variables (past returns), forecasting models (nonlinear models), and alternative forecasting evaluation criteria (economic criteria). Our analysis suggests that (growth) stock returns might be predictable.

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

JEL Classifications: C2, C5, G15

**Corresponding author*: The Business School, PO Box 173364, University of Colorado Denver Denver, CO 80217-3364. Email: Jian.Yang@ucdenver.edu; Tel: (303) 556-5852; Fax: (303) 556-5899.

We would like to thank Yongmiao Hong and Tae-Hwy Lee for sharing some computer programs used in this study; Ken French for making the data available through his website; Donald R. Fraser, Hui Guo, Dennis W. Jansen, Qi Li, Robert A. Van Ness and seminar/session participants at Texas A&M University and the 2009 Eastern Finance Association for much help and suggestions. We are grateful to the editor (Arnie Cowan) and the anonymous referee for very extensive comments, which greatly improve the paper. Wang and Yang acknowledge financial support from the PSC-CUNY research grant and the UC Denver Faculty Research & Scholarship Completion Fund, respectively.

1. Introduction

Whether stock returns are predictable has been one of the most active research topics in finance, especially since Fama's (1965) seminal work on the efficient market hypothesis. However, most relevant to this study, the majority of earlier works on testing the predictability of stock returns based on past returns fail to reject the martingale hypothesis. For example, many studies employing the variance ratio test of Lo and MacKinlay (1988) at best yield a mixed conclusion, especially for the low-frequency,(e.g., monthly) data (e.g., Ayadi and Pyun, 1994; Urrutia, 1995; Coggin, 1998; Abraham, Seyyed, and Alsakran, 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 it is difficult for most financial economists and professionals to predict asset prices. However, both the popular autocorrelation and variance ratio tests used in the literature assume linearity and only test serial uncorrelatedness rather than the martingale difference (e.g., Hsieh, 1991; McQueen and Thorley, 1991; Hong and Lee, 2003). Because a nonlinear time series can have zero autocorrelation and a non-zero mean conditional on its past history (i.e., predictable based on the past history), both tests can fail to capture predictable nonlinearities in the mean and could yield misleading conclusions in favor of the martingale hypothesis.

Using a comprehensive set of nonlinear models, we reexamine international stock return predictability based on past returns, which is at the heart of testing weak-form market efficiency. As Grinblatt and Moskowitz (2004) point out, past returns contain information about expected returns, which is readily available to be exploited. This 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 of the most important equity style indexes (growth and value stock indexes) for thirteen major international stock markets over the

period January 1975 to December 2004. Testing the martingale behavior of style equity indexes is important in itself because style investing has become popular in the last two decades, and its performance benchmarks are style indexes (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; Zhang, 2005) models. Surprisingly, with the notable exception of Coggin (1998), few authors have addressed the predictability of growth and value portfolios based on their past price information. We fill this gap by providing a comprehensive investigation of the predictability of style indexes.

Second, we employ the model selection approach (e.g., Swanson and White, 1997) and comprehensively address stock return predictability in a nonlinear, out-of-sample context. Compared to the traditional hypothesis testing approach that 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 in making the correct inference based on nonparametric models because they notoriously tend to overfit the data. Goyal and Welch (2008) recently underscore the importance of out-of-sample forecasting performance in understanding stock return predictability. Second, unlike the traditional hypothesis testing approach, the model selection approach 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 as identified by Granger (1992): 1) benefits can arise, especially when considering non-linear models and 2) only out-of-sample evaluation is relevant and (to some extent) avoids data mining difficulties. Furthermore, to address potential data snooping biases,

we conduct White's (2000) test for out-of-sample multiple model comparison. White (2000) constructs a method for testing the hypothesis that the best model encountered during a specification search has no predictive superiority over the benchmark model. In the hypothesis testing, the model permits data snooping, with some degree of confidence that no one will mistake the results generated by chance for genuinely "good" results.

A few earlier works have addressed the predictability of international stock market returns using the variance ratio test (e.g., Patro and Wu, 2004) or other economic variables (e.g., Guo, 2006). However, the authors often only focus on in-sample evidence and, more importantly, do not allow for potential predictable nonlinearity-in-mean. Also noteworthy, many studies (e.g., Leitch and Tanner, 1991; Brock, Lakonishok, and LeBaron, 1992; Granger, 1992; Gencay, 1998, 1999) have emphasized the importance of trading rule profitability (as an economic criterion) to evaluate forecasting performance. However, the direction of changes as an alternative economic criterion has not really been much explored. From the perspective of decision making under uncertainty, there exists important circumstances under which this criterion is exactly the right one for maximizing the economic welfare of the forecaster (e.g., Leitch and Tanner, 1991; Hong and Lee, 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 perhaps 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.

2. Econometric methods

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 moments, and its conditional mean, $E(Y_t \mid I_{t-1})$ may be time-varying in a complicated nonlinear form.[1] Thus, we will include various popular nonlinear parametric and nonparametric models used in the literature.[2] Barnett and Serletis (2000) provide a rather comprehensive literature review on the role of nonlinearity in testing the martingale hypothesis. McQueen and Thorley (1991) suggest that nonlinear patterns in stock returns might arise because of fads or rational speculative bubbles. Hsieh (1991) argues 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. Urrutia, Vu, Gronewoller, and Hoque (2002) also demonstrate how apparent randomness of stock price movement can be explained by nonlinear dynamic systems.

We use the martingale model (with a drift) as the benchmark model, and we consider the following four popular nonlinear models as comparison: the polynomial regression model (PN), the artificial neural network (NN), the functional coefficient model (FC), and the nonparametric regression model (NP). We also consider a basic linear model, the autoregression model [i.e., the AR (1)] model, which is a popular null model for modeling stock returns (e.g., Brock, Lakonishok, and LeBaron, 1992). Certainly, 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

[1] The terms "random walk" and "martingale" have been interchangeably used in the efficient capital markets literature. However, it is the martingale property (or unpredictability) of security prices that is of essential interest to this huge body of the literature (Fama, 1965).

[2] This distinguishes our study from some recent studies (e.g., Al-Khazali, Ding, and Pyun, 2007) using the nonparametric variance ratio test. Our study also further differs in using several (rather than one) nonparametric nonlinear techniques and exploring economic criteria (rather than statistical criteria only) (e.g., Darrat and Zhong, 2000).

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. 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 neural networks combines many 'basic' nonlinear functions via a multilayer structure. Normally there is one intermediate (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:

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

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

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. The correct number of lags needed is typically unknown, and in some instances lagged dependent variables may not be sufficient in capturing 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 the input variables are weighted (differently) and are 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 choose 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 (FC) model, first proposed by Cai, Fan, and Yao (2000), is a newer nonlinear time series model with state-dependent coefficients, including 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, that can 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 $t-1$; also stated as follows:

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

We use $L = 12$ to capture the 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 for forecasting the direction of stock price changes.

Similar to Cai, Fan, and Yao (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 we use h as 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, Fan, and Yao (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 can 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 and simply using 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 $h = (h_1, \ldots h_j)$ is chosen by the leave-one-out cross-validation procedure. As already noted, smoothing parameter h is the most important parameter in nonparametric estimation. An inappropriately chosen h will give poor in-sample and out-of-sample predictions. 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.

Our modified approach consists of finding the best h for out-of-sample forecasting and making forecasts based on this h^*.[3] In this procedure, we have two parameters to establish in

[3] For example, suppose that we have data points x_1 to x_{100} and we want to forecast x_{101}. We propose the following modified nonparametric forecasting method. We use h^* and data points x_1 to x_{80} to forecast x_{81}, …, data points

the example: (1) the out-of-sample evaluation length k is set equal to 20 ($\hat{x}_{81}$ to $\hat{x}_{100}$), and (2) the regression length m is set equal to 80. 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 can have different time series properties. For a more (less) volatile time series, a shorter (longer) evaluation length (m) may be better and vice versa. Thus, we choose several 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 the impact is not very substantial qualitatively for the cases considered in this study.

2.4. Evaluation criteria

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 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

---

x_{20} to x_{99} 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}.

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).[4] 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.

3. Data and results

3.1. Data descriptions

Monthly return data for this study cover a 30-year period from January 1975 to December 2004 for 13 major international stock markets (except for Canada, where the data start from January 1977): Australia (AU), Belgium (BE), Canada (CA), France (FR), Germany (GE), 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 obtain the data of the market, growth, and value indexes for each country from the website of Ken French. Value (growth) portfolios consist of firms where the book-to-market ratio is among the highest (lowest) 30th 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 denoted in the local currency.

[4] Closely following Fama (1991) and Gencay (1998, 1999), we do not explicitly allow for transaction costs in the evaluation of trading rule performance for various models. According to Fama (1991), the researcher 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).

We focus on monthly return data because international value and growth return data are only available to us at the monthly frequency. Monthly data are actually more appropriate for the purpose of this paper than daily or weekly data because 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; Hsieh, 1991). Therefore, monthly data should provide cleaner evidence of stock return predictability than daily or weekly data.

3.2. The results on stock market indexes

The goal of this paper is rather modest: we are primarily interested in whether past stock returns can be useful in forecasting future stock returns when nonlinear models and economic criteria are also considered, which have received inadequate attention in this line of the literature. Because the benchmark is always the martingale model, there is some evidence in favor of stock return predictability, as long as some models can outperform the benchmark in certain criteria. The model selection approach is agnostic in its spirit and does not require determining whether a particular model, which may be practically termed as the correct model, may apply well to all cases under examination. Indeed, it would be surprising if one particular model would outperform the benchmark in all markets because different markets can be characterized by somewhat different economic dynamic systems. More specifically, different economies might have different responses to more or less different economic forces, and stock returns may not be similarly related to the economic forces. Hence, using past returns as predictors, we should not assume that a particular (linear or nonlinear) dynamic model would prevail across all or even most markets.[5]

[5] We thank the referee for bringing up this point.

We use the rolling regression technique to make out-of-sample forecasts. The rolling regression 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. Swanson and White (1997) argue that the rolling regression technique is advantageous because it further allows for the (potentially nonlinear) relation between the current and past returns to evolve across time. Specifically, suppose that there are N observations in total, where $N = R + P$, and P is the number of observations reserved for measuring the out-of-sample forecast performance. When the rolling regression technique is applied, we use the first R observations for the in-sample estimation and then generate the one-step-ahead forecast for period $R+1$, which we can denote $\hat{P}_1$. Then we use observations from period two to period $R+1$, which again correspond to the fixed-length window of R observations, for the in-sample estimation and then generate the one-step-ahead forecast for period $R+2$ or $\hat{P}_2$. By repeating for a total of P times, we generate a sequence of out-of-sample one-step-ahead return forecasts $\hat{P}_1, \hat{P}_2 \ldots \hat{P}_P$, and we can compare them with actual values of the P observations saved for measuring the out-of-sample forecast performance, which would generate a sequence of one-step-ahead forecast errors.

The one-step-ahead forecast errors form the basis for the calculation of MSFE and MAFE, and the out-of-sample one-step-ahead return forecasts $\hat{P}_1, \hat{P}_2 \ldots \hat{P}_P$ are the only required input for determining MFTR and MCFD. Also, since all the evaluation criteria are based on the mean or average performance of forecast errors, trading returns, and correct forecast direction over the whole out-of-sample period, the consistency of forecasting performance for various models has to some extent been addressed. Finally, different from many earlier studies, we do not assume an ad hoc distribution when we attempt to determine whether the MSFE, MAFE, MFTR, or MCFD from forecasting models (which are all the averages of the corresponding measures) are

statistically different from those of the benchmark. Instead, we use the bootstrapping technique to obtain the empirical distributions that exactly match the data and thus the corresponding *p*-values.

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 a ratio with 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 the out-of-sample forecast. Nevertheless, for robustness, we also experiment with alternative ratios of $R:P=3:1$ and $R:P=4:1$.

Tables 1 and 2 report the out-of-sample forecast results for thirteen international stock market indexes. All the 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 ratios relative to that of the benchmark model for the other forecasting models. The vast majority of the MSFE and MAFE ratios are greater than

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 based on MSFE (both significant at the 10% level) and the NN model for Singapore based on MAFE (significant at the 5% level). Therefore, consistent with previous studies (e.g., Hsieh, 1991), the forecasting models do not outperform the benchmark model based on statistical criteria. Also, the relative usefulness of the NN model when compared to other nonparametric models for forecasting the Singapore market is somewhat consistent with Gencay (1998). Nevertheless, although Gencay (1998) shows some evidence of significant nonlinear forecasting performance based on MSFE for the daily Dow Jones Industrial Average indexes, we do not find such evidence for the monthly CRSP value-weighted index and our finding is consistent with Hsieh (1991). The difference could be attributed to the pronounced nonsynchronous trading problem of the daily price index, differences in the broad-basedness of the indexes, and differences 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 based on MFTR at the 5% and 10% significance levels, respectively. 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, based on MCFD, in addition 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 when 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 change 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 possibly Belgium, Canada, Japan, and Singapore, major international stock markets appear to be unpredictable.[6] We might attribute the predictability of Belgium, Canada, and Singapore to their relatively small market size, and a possible explanation for Japan might be its overly prolonged economic slowdown and downward or stagnant stock market movement since the early 1990s.

The predictability for stock market indexes is also often detected by the NN model. This result is consistent with widely perceived usefulness of artificial neural network in uncovering the nonlinearity-in-mean. It is also noteworthy that none of the 13 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 can increase the power for uncovering the 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.) based on statistical criteria (Table 3). Specifically, although the evidence for Canada is significant only at the 10% level based on either MSFE or MAFE criterion, the forecast improvements can be substantial (with the MSFE ratio of 0.91 based on the NP model). The predictability evidence for Hong Kong is significant at the 5% level based on MSFE but only at the 10% level based on

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 ratios of 0.99). The most statistically significant evidence exists for the U.S. value stock portfolio. The evidence is significant at the 5% level based on 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 change 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) of improved predictions of price-change direction 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 based on both MFTR and MCFD. While we find that the general stock market for Singapore is predictable based on statistical criteria, there is no such evidence for its value stock portfolios. Nevertheless, the predictability evidence for the country shows up based on 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

[6] Hjalmarsson (2006) shows that when testing for stock return predictability, the sample sizes in most relevant cases are simply too small to obtain precisely estimated predictive relations that would show up in out-of-sample

dominates (0.83% per month) that of the benchmark model.

To summarize, based on the statistical criteria, there is evidence against the martingale hypothesis for value stock portfolios in three of 13 international stock markets, namely, Canada, Hong Kong, and the U.S; based on the economic criteria, there is evidence in four stock markets, namely, Belgium, Hong Kong, Singapore, and the U.S. These five countries include three of the four countries where the general stock markets are found to be predictable (with the only exception of Japan, for which there is also some predictability evidence based on the MCFD significant at the 15% level).

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 based on statistical criteria (either MSFE or MAFE), with the exception of Belgium, for which both the AR and PN models outperform the benchmark model based on 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 eight countries. This result clearly demonstrates the importance of considering economic criteria in the forecast evaluation, as Hong and Lee (2003) and Yang, Su, and Kolari (2008) emphasize 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 can fail to detect the superiority of the nonlinear models.[7]

Specifically, there is predictability evidence based on MFTR in Table 6—all of which are

exercises. In this context, our evidence of the out-of-sample predictability could be a conservative estimate.

[7] As pointed out by the referee, the reason why economic criteria may produce better evidence in this study could be due to the fact that economic criteria impose less burden on the correlation structure of the relations between dependent and independent variables than statistical criteria, since the correlation structure is only directional and not in levels.

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-change direction can be improved over the benchmark model for Canada, Hong Kong, Switzerland, and the U.K., which is significant at the 10% level. In addition, we find new predictability evidence for France at the 5% level.

To summarize, the forecasting models outperform the benchmark model for growth stock portfolios in nine out of the thirteen countries, including Belgium, Canada, France, Hong Kong, Japan, Netherlands, Singapore, Switzerland, and the U.K. These nine countries include all four of the countries where the general stock markets show the evidence of predictability and four of the five countries where the value-stock portfolios show evidence of predictability.[8]

It is somewhat puzzling that growth stocks are not predictable in the U.S. but are predictable in the majority of 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 (2008) use theoretically motivated financial variables (i.e., stock market volatility and average idiosyncratic volatility) to forecast stock returns, and they find that the in-sample R-squared is substantially higher for growth stocks than value stocks. Therefore, our international evidence can be consistent with the finding of Guo and Savickas (2008). Another possible explanation could be that the U.S. market for growth stocks is more efficient than other markets, and thus it is more difficult to be predicted. Consistent with this conjecture, we did not find evidence of predictability for the U.S. general stock market. Also, unlike the case for stock market indexes,

[8] The list of the countries with the evidence of predictability shows some sensitivity when evaluated with alternative statistical criteria such as AIC or SIC. In particular, Belgium may not possess statistically significant predictability

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), which is the major advantage of the model selection approach.

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 from the authors. 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. 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, based on 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

anymore. Nevertheless, the results based on economic criteria remain unchanged and yield the similar conclusion.

(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 $R:P=4:1$, there is still much evidence of predictability, particularly for growth stocks.

Patro and Wu (2004) also conduct a 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, the results are not tabulated here but are available from the authors.

To further 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. The majority of significant test statistics based on a single model remains qualitatively the same. This statement is particularly true, given the concerns of a 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 evidence of return predictability, particularly for growth stocks, 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. In particular, we investigate two subsamples: July 1926 to December 1970 (with a total of 534 observations) and January 1970 to December 2004 (with a total of 408 observations). Table 7 shows that, based on the ratio $R:P=2:1$, the market index, the value stock portfolio, and the growth stock portfolio are all predictable by the AR model, based on economic criteria but not statistical criteria. The PN model also detects predictability in the value stock index at the 10% significance level. However, this finding is somewhat sensitive

to the alternative ratio of $R:P=4:1$, for which we find none of the three indexes as predictable at the 10% level although two of them are predictable at the 15% level.

For the second subsample, the results (available from the authors) 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, while there is little predictability for the market index and growth stock index. Thus, the additional results for the second subsample confirm that our main finding based on a shorter 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 predictability 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 much evidence of predictability for growth stocks, especially when evaluated with economic criteria and to a noticeably lesser extent for the general stock markets and value stocks. To the best of our knowledge, the documented short-horizon predictability pattern based on international data is novel in this line of the literature.

Our results shed light on the recent debate about stock return predictability. Goyal and Welch (2008) reexamine the existing empirical evidence and find little support of the out-of-sample predictability. We address this issue using different assets (growth style indexes), forecasting variables (past returns), forecasting models (nonlinear models), and alternative forecasting evaluation criteria (economic criteria). Our analysis suggests that (growth) stock

returns might be predictable.

Finally, consistent with previous studies (e.g., Leitch and Tanner, 1991; Hong and Lee, 2003; Yang, Su, and Kolari, 2008), 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 eight countries.

References

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

Al-Khazali, O.M., D.K. Ding, and C.S. Pyun, 2007. A new variance ratio test of random walk in emerging markets: A revisit, *The Financial Review* 42, 303-317.

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

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

Barnett, W.A. and A. Serletis, 2000. Martingales, nonlinearity, and chaos, *Journal of Economic Dynamics and Control* 24, 703-724.

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

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

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

Campbell, J. and T. Vuolteenaho, 2004. Bad beta, good beta, *American Economic Review* 94, 1249-1275.

Chaudhuri, K. and Y. Wu, 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 J. Smith, 2001. Evaluating forecasts from SETAR models of exchange rates, *Journal of International Money and Finance* 20, 133-148.

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

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

Darrat, A. and M. Zhong, 2000. On testing the random walk hypothesis: A model-comparison approach, *The 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 I. Welch, 2008. A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455-1508.

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

Guo, H. and R. Savickas, 2008. Aggregate idiosyncratic volatility in G7 countries, *Review of Financial Studies* 21, 1259-1296.

Guo, H., 2006. On the risk-return relation in international stock markets, *The Financial Review* 41, 565-587

Grinblatt, M. and T.J. Moskowitz, 2004. Predicting stock price movements from past returns: The role of consistency and tax-loss selling, *Journal of Financial Economics* 71, 541-579.

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

Hjalmarsson, E., 2006. Should we expect significant out-of-sample results when predicting stock returns? International Finance discussion paper # 855, Board of Governors of the Federal Reserve.

Hong, Y. M. and T.H. Lee, 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.

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

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

Malkiel, B.G., 2005. Reflections on the efficient market hypothesis: 30 years later, *The Financial Review* 40, 1-9.

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

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

Swanson, N.R. and H. White, 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.

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., J. Vu, P. Gronewoller, and M. Hoque, 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.

Yang, J., X. Su, and J.W. Kolari, 2008. Do Euro exchange rates follow a martingale? Some out-of-sample evidence, *Journal of Banking and Finance* 32, 729 - 740.

Zhang, L, 2005. The value premium, *Journal of Finance* 60, 67-103.

Table 1
Forecast evaluation for stock market indexes – statistical criteria

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. AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. The numbers in the parentheses are bootstrapped *p*-values of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

	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
	(0.88)	(0.33)	(0.09)	(0.52)	(0.73)	(0.49)	(0.91)	(0.28)	(0.66)	(0.47)	(0.29)	(0.78)	(0.77)
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
	(0.82)	(0.61)	(0.93)	(0.55)	(0.90)	(0.75)	(0.96)	(0.88)	(0.98)	(0.87)	(0.34)	(0.94)	(0.86)
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
	(0.80)	(0.18)	(0.93)	(0.94)	(0.71)	(0.90)	(0.29)	(0.87)	(0.45)	(0.08)	(0.75)	(0.98)	(0.63)
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
	(0.96)	(0.95)	(0.93)	(0.77)	(0.93)	(0.72)	(0.77)	(0.87)	(0.92)	(0.90)	(0.89)	(0.80)	(0.95)
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
	(0.86)	(0.54)	(0.96)	(0.87)	(0.80)	(0.55)	(0.93)	(0.34)	(0.89)	(0.62)	(0.12)	(0.88)	(0.82)
	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
	(0.74)	(0.15)	(0.17)	(0.35)	(0.74)	(0.70)	(0.96)	(0.21)	(0.90)	(0.97)	(0.40)	(0.94)	(0.75)
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
	(0.94)	(0.16)	(0.85)	(0.82)	(0.88)	(0.59)	(0.95)	(0.72)	(0.97)	(0.93)	(0.52)	(0.93)	(0.69)
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
	(0.59)	(0.40)	(0.97)	(0.94)	(0.76)	(0.99)	(0.78)	(0.60)	(0.61)	(0.04)	(0.65)	(0.97)	(0.52)
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
	(0.92)	(0.94)	(0.85)	(0.80)	(0.93)	(0.80)	(0.62)	(0.76)	(0.97)	(0.97)	(0.85)	(0.87)	(0.97)
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
	(0.99)	(0.46)	(0.91)	(0.82)	(0.85)	(0.56)	(0.86)	(0.23)	(0.86)	(0.73)	(0.14)	(0.67)	(0.54)

Table 2
Forecast evaluation for stock market indexes – economic criteria

The monthly data cover the period January 1975 to December 2004. The benchmark model is the martingale model. AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. The numbers in the parentheses are bootstrapped *p*-values of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

	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)

Table 3
Forecast evaluation for value stock indexes – statistical criteria

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. AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. The numbers in the parentheses are bootstrapped *p*-values of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

	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
	(0.47)	(0.42)	(0.73)	(0.55)	(0.94)	(0.02)	(0.68)	(0.65)	(0.40)	(0.35)	(0.15)	(0.59)	(0.22)
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
	(0.97)	(0.77)	(0.20)	(0.79)	(0.91)	(0.51)	(0.92)	(0.86)	(0.71)	(0.79)	(0.61)	(0.83)	(0.89)
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
	(0.15)	(0.24)	(0.80)	(0.93)	(0.91)	(0.36)	(0.83)	(0.73)	(0.93)	(0.20)	(0.72)	(0.99)	(0.03)
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
	(0.97)	(0.76)	(0.97)	(0.89)	(0.92)	(0.89)	(0.83)	(0.92)	(0.94)	(0.89)	(0.96)	(0.97)	(0.73)
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
	(0.76)	(0.85)	(0.06)	(0.86)	(0.89)	(0.17)	(0.88)	(0.95)	(0.79)	(0.89)	(0.18)	(0.89)	(0.82)
	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
	(0.37)	(0.52)	(0.92)	(0.88)	(0.85)	(0.09)	(0.89)	(0.31)	(0.25)	(0.60)	(0.28)	(0.51)	(0.28)
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
	(0.99)	(0.71)	(0.37)	(0.95)	(0.94)	(0.48)	(0.96)	(0.83)	(0.60)	(0.93)	(0.57)	(0.82)	(0.60)
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
	(0.20)	(0.44)	(0.86)	(0.98)	(0.96)	(0.80)	(0.98)	(0.36)	(0.96)	(0.72)	(0.63)	(0.92)	(0.04)
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
	(0.99)	(0.87)	(0.99)	(0.94)	(0.88)	(0.90)	(0.76)	(0.89)	(0.90)	(0.93)	(0.99)	(0.98)	(0.83)
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
	(0.70)	(0.79)	(0.06)	(0.92)	(0.93)	(0.32)	(0.97)	(0.89)	(0.73)	(0.96)	(0.18)	(0.87)	(0.63)

Table 4
Forecast evaluation for value stock indexes – economic criteria

The monthly data cover the period January 1975 to December 2004. The benchmark model is the martingale model. AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. The numbers in the parentheses are bootstrapped *p*-values of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

	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
	(0.96)	(0.88)	(0.89)	(0.33)	(0.96)	(0.79)	(0.35)	(0.74)	(0.43)	(0.67)	(0.17)	(0.74)	(0.72)
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
	(0.94)	(0.95)	(0.91)	(0.81)	(0.52)	(0.75)	(0.52)	(0.67)	(0.38)	(0.79)	(0.25)	(0.73)	(0.75)
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
	(0.74)	(0.71)	(0.33)	(0.97)	(0.84)	(0.78)	(0.96)	(0.79)	(0.76)	(0.06)	(0.55)	(0.94)	(0.08)
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
	(0.74)	(0.57)	(0.97)	(0.80)	(0.96)	(0.61)	(0.32)	(0.94)	(0.28)	(0.90)	(0.28)	(0.91)	(0.65)
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
	(0.75)	(0.08)	(0.25)	(0.66)	(0.16)	(0.27)	(0.65)	(0.91)	(0.47)	(0.81)	(0.66)	(0.62)	(0.96)
	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
	(0.96)	(0.85)	(0.87)	(0.31)	(0.96)	(0.60)	(0.65)	(0.74)	(0.59)	(0.75)	(0.25)	(0.75)	(0.74)
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
	(0.94)	(0.96)	(0.93)	(0.79)	(0.45)	(0.44)	(0.56)	(0.12)	(0.41)	(0.86)	(0.23)	(0.75)	(0.74)
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
	(0.73)	(0.65)	(0.44)	(0.92)	(0.95)	(0.37)	(0.86)	(0.61)	(0.91)	(0.21)	(0.37)	(0.97)	(0.07)
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
	(0.74)	(0.85)	(0.99)	(0.77)	(0.98)	(0.06)	(0.29)	(0.92)	(0.24)	(0.84)	(0.36)	(0.76)	(0.93)
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
	(0.61)	(0.08)	(0.16)	(0.56)	(0.22)	(0.43)	(0.81)	(0.65)	(0.62)	(0.73)	(0.57)	(0.96)	(0.17)

Table 5
Forecast evaluation for growth stock indexes – statistical criteria

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. AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. The numbers in the parentheses are bootstrapped *p*-values of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

	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
	(0.93)	(0.11)	(0.81)	(0.30)	(0.74)	(0.57)	(0.96)	(0.73)	(0.93)	(0.83)	(0.17)	(0.75)	(0.79)
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
	(0.93)	(0.24)	(0.85)	(0.20)	(0.93)	(0.76)	(0.89)	(0.81)	(0.71)	(0.84)	(0.31)	(0.99)	(0.82)
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
	(0.98)	(0.33)	(0.92)	(0.43)	(0.70)	(0.29)	(0.99)	(0.89)	(0.11)	(0.37)	(0.84)	(0.99)	(0.62)
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
	(0.99)	(0.83)	(0.97)	(0.76)	(0.97)	(0.54)	(0.71)	(0.91)	(0.97)	(0.97)	(0.87)	(0.94)	(0.91)
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
	(0.78)	(0.83)	(0.93)	(0.49)	(0.83)	(0.92)	(0.88)	(0.61)	(0.70)	(0.82)	(0.56)	(0.88)	(0.88)
	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
	(0.91)	(0.04)	(0.20)	(0.16)	(0.50)	(0.82)	(0.98)	(0.81)	(0.98)	(0.94)	(0.17)	(0.93)	(0.92)
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
	(0.98)	(0.03)	(0.81)	(0.29)	(0.89)	(0.50)	(0.86)	(0.79)	(0.87)	(0.83)	(0.36)	(0.99)	(0.95)
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
	(0.99)	(0.38)	(0.75)	(0.65)	(0.75)	(0.53)	(0.99)	(0.80)	(0.27)	(0.89)	(0.67)	(0.99)	(0.71)
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
	(0.99)	(0.85)	(0.99)	(0.68)	(0.95)	(0.62)	(0.88)	(0.91)	(0.99)	(0.98)	(0.87)	(0.96)	(0.91)
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
	(0.91)	(0.78)	(0.88)	(0.47)	(0.76)	(0.94)	(0.51)	(0.27)	(0.95)	(0.89)	(0.66)	(0.87)	(0.83)

Table 6
Forecast evaluation for growth stock indexes – economic criteria

The monthly data cover the period January 1975 to December 2004. The benchmark model is the martingale model. AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. The numbers in the parentheses are bootstrapped *p*-values of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

	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
	(0.96)	(0.11)	(0.08)	(0.75)	(0.90)	(0.96)	(0.96)	(0.41)	(0.96)	(0.96)	(0.59)	(0.97)	(0.96)
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
	(0.74)	(0.12)	(0.23)	(0.75)	(0.87)	(0.76)	(0.94)	(0.43)	(0.74)	(0.87)	(0.09)	(0.96)	(0.94)
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
	(0.79)	(0.47)	(0.61)	(0.78)	(0.22)	(0.22)	(0.90)	(0.25)	(0.08)	(0.10)	(0.57)	(0.93)	(0.53)
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
	(0.96)	(0.47)	(0.25)	(0.44)	(0.91)	(0.96)	(0.50)	(0.10)	(0.30)	(0.44)	(0.81)	(0.84)	(0.89)
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
	(0.84)	(0.86)	(0.82)	(0.15)	(0.25)	(0.84)	(0.72)	(0.06)	(0.74)	(0.52)	(0.66)	(0.08)	(0.74)
	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
	(0.91)	(0.11)	(0.09)	(0.73)	(0.65)	(0.96)	(0.94)	(0.57)	(0.96)	(0.93)	(0.61)	(0.96)	(0.94)
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
	(0.73)	(0.19)	(0.33)	(0.65)	(0.56)	(0.34)	(0.96)	(0.74)	(0.74)	(0.81)	(0.09)	(0.92)	(0.96)
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
	(0.48)	(0.57)	(0.59)	(0.82)	(0.31)	(0.09)	(0.68)	(0.40)	(0.20)	(0.52)	(0.42)	(0.92)	(0.70)
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
	(0.86)	(0.58)	(0.22)	(0.12)	(0.73)	(0.94)	(0.84)	(0.21)	(0.69)	(0.47)	(0.69)	(0.58)	(0.95)
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
	(0.86)	(0.72)	(0.74)	(0.05)	(0.39)	(0.81)	(0.33)	(0.19)	(0.74)	(0.68)	(0.63)	(0.08)	(0.42)

Table 7

Forecast evaluation for the U.S.: July 1926 to December 1970 and R:P = 2:1

The actual values of MSFE and MAFE are reported for the benchmark model (i.e., the martingale). AR, PN, NN, FC, and NP stand for the autoregressive, polynomial, artificial neural network, functional coefficient, and nonparametric regression models, respectively. 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 of White's (2000) Reality Check test statistics for testing no difference in the forecasting performance between each of the models and the benchmark.

(R,P) = (347,173)		MSFE		MAFE		MFTR		MCFD	
k	Model	MSFE	*P*	MAFE	*P*	MFTR	*P*	MCFD	*P*
		Panel A. The market index							
0	Benchmark	13.95		2.92		0.73		0.64	
1	AR	0.99	0.36	1.00	0.68	0.78	0.08	0.65	0.07
2	PN	0.98	0.22	1.00	0.54	0.67	0.67	0.64	0.59
3	NN	1.06	0.93	1.02	0.82	0.63	0.75	0.64	0.48
4	FC	1.02	0.70	1.02	0.76	0.55	0.85	0.62	0.87
5	NP	1.05	0.87	1.02	0.88	0.69	0.66	0.64	0.65
		Panel B. The value stock index							
0	Benchmark	18.54		3.36		1.08		0.65	
1	AR	0.98	0.15	0.98	0.01	1.10	0.44	0.67	0.08
2	PN	0.99	0.29	1.00	0.36	1.08	0.08	0.67	0.08
3	NN	0.99	0.35	1.00	0.24	1.08	0.99	0.65	0.99
4	FC	1.07	0.88	1.08	0.87	0.77	0.92	0.65	0.74
5	NP	1.00	0.51	1.00	0.42	1.10	0.37	0.64	0.35
		Panel C. The growth stock index							
0	Benchmark	15.50		3.09		0.66		0.59	
1	AR	0.99	0.33	1.00	0.62	0.71	0.08	0.60	0.09
2	PN	0.99	0.23	1.01	0.66	0.72	0.38	0.58	0.68
3	NN	1.06	0.93	1.04	0.96	0.50	0.82	0.57	0.77
4	FC	1.39	0.89	1.12	0.88	0.39	0.84	0.57	0.82
5	NP	1.08	0.87	1.02	0.78	0.54	0.82	0.59	0.78