Are the Directions of Stock Price Changes Predictable? Statistical Theory and Evidence

Yongmiao Hong and Jaehun Chung
Cornell University

March 2003

Correspondence: Yongmiao Hong, Department of Economics & Department of Statistical Science, Uris Hall, Ithaca, NY 14850, USA, and Department of Economics, School of Economics and Management, Tsinghua University. Email: yh20@cornell.edu, Tel: 1-607-255 5130, Fax: 1-607-255 2818. Jaehun Chung, Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14850, USA. Email: jc257@cornell.edu, Tel: 1-607-255 4254, fax: 1-607-255 2818. We thank the National Science Foundation for financial support via NSF Grant SES-0111769. Any remaining errors are solely ours.

Abstract

We propose a model-free omnibus statistical procedure to check whether the direction of changes in an economic variable is predictable using the history of its past changes. A class of separate inference procedures are also given to gauge possible sources of directional predictability. They can reveal information about whether the direction of future changes is predictable using the direction, level, volatility, skewness, and kurtosis of past changes. An important feature of the proposed procedures is that they check many lags simultaneously, which is particularly suitable for detecting the alternatives whose directional dependence is small at each lag but it carries over a long distributional lag. At the same time, the tests naturally discount higher order lags, which is consistent with the conventional wisdom that financial markets are more influenced by the recent past events than by the remote past events.

We apply the proposed procedures to four daily U.S. stock price indices. We find overwhelming evidence that the directions of excess stock returns are predictable using past excess stock returns, and the evidence is stronger for the directional predictability of large excess stock returns. In particular, the direction and level of past excess stock returns can be used to predict the direction of future excess stock returns with any threshold, and the volatility, skewness and kurtosis of past excess stock returns can be used to predict the direction of future excess stock returns with nonzero thresholds (i.e., large returns). The well-known strong volatility clustering together with weak serial dependence in mean cannot completely explain all documented directional predictability for stock returns. To exploit the economic significance of the documented directional predictability for stock returns, we consider a class of autologit models for directional forecasts and find that they have significant out-of-sample directional predictive power. Some trading strategies based on these models and their combinations can earn significant out-of-sample extra risk-adjusted returns over the buy-and-hold trading strategy.

Key words: Autologit models, Characteristic function, Combined forecasts, Directional predictability, Efficient market hypothesis, Generalized spectrum, Market timing, Sharpe Ratio, Volatility clustering.
JEL NO: C12, C22, G0

1 Introduction

Predictability of asset returns has immediate interest for investment practitioners and far-reaching implications for the efficacy of asset prices in allocating capitals. Focus in this literature has been on the predictability of the level or conditional mean of asset returns (e.g., Fama 1970, 1991, Jegadeesh 1990, Lo and MacKinlay 1999, Poterba and Summers 1988). In this paper, we investigate the predictability of the direction of changes in economic variables, such as interest rates, inflation rates, exchange rates and stock prices. The direction of changes in economic variables may be a reasonable proxy for a utility-based measure of forecasting performance. Leitch and Tanner (1991, 1995) find that the direction-of-change criterion is the best proxy among several commonly used criteria for choosing forecasts of interest rates on their ability to maximize expected trading profits. There exist important circumstances under which the direction-of-change criterion is exactly the right one for maximizing the welfare (e.g., profit) of the forecaster, as is nicely demonstrated in Granger and Pesaran (1999, Sections 2-4) and Leitch and Tanner (1995) from a perspective of decision-making under uncertainty. Macroeconomists and investment practitioners have been interested in forecasting probabilities of important economic events (e.g., Diebold and Lopez 1996, Fair 1993), which, in many cases, can be formulated as the probabilities of the direction of changes in underlying economic variables. Central banks under pegged exchange rate systems, for example, are often interested only in the direction of changes in the exchange rate. They might need to intervene to support the currency if it is expected to depreciate, regardless of the size of the expected depreciation. Over the past few years, some central banks, including the Bank of England, have been setting the nominal interest rate according to their forecasts of the inflation rate, increasing the interest rate if their forecast of the inflation rate exceeds a politically determined threshold. In finance, directional predictability in asset returns has important implications for market timing, which is crucial for active asset allocation management. In Merton's (1981) classical market timing model, mutual fund managers care about the direction of excess returns, rather than their magnitude. Most commonly used technical trading rules in financial markets are based on the prediction of the directions of financial returns. Profitable trading strategies may result if one can predict return directions. Many financial institutions evaluate forecast algorithms using the percentage of times that the algorithms predict the right-trend (see Lequarre 1993).

The rationale behind directional forecasts is that the patterns in economic variables may recur in the future so that the direction of changes in economic variables is predictable using historical data. The main goal of this paper is to develop a mode-free omnibus test for directional predictability and apply it to document whether the direction of stock price changes

is predictable using the history of past stock price changes. Most, if not all, of the existing works in this literature are concerned with directional predictability of various models, algorithms, and investment strategies. There have been a number of popular tests for the market timing ability of these models and trading strategies (Henriksson and Merton 1981, Cumby and Modest 1987, Pesaran and Timmermann 1992). However, the directional predictability of an underlying data generating process is not the same as the predictive ability of a directional forecast model or a trading strategy. There has been no model-free test available in the literature that can check directional predictability of the data, which is the key to the success of any directional forecast model or trading strategy.

Some economic and financial theory suggests that the direction of asset returns may be predictable. For example, the naive overreaction theory predicts price reversals after investors overreact to certain market events such as release of firm-specific information, which implies a negative autocorrelation in direction. More sophisticated behavioral theory (e.g., Barberis, Sheleifer and Vishny 1998, Hong and Stein 1999) predicts a short-horizon underreaction and then a long horizon overreaction, implying positive autocorrelations in direction over a short horizon and negative autocorrelations in direction over a long horizon. The market contagion hypothesis, on the other hand, suggests that during a turmoil period, a large adverse price movement in one market will be more closely followed by a large adverse price movement in another market, regardless of market fundamentals. This implies a stronger positive cross-correlation in direction between two markets during the turmoil period.[1] In the foreign exchange markets, it is often argued that the exchange rate may follow long swings — it drifts upward for a considerable period of time and then switches to a long period with downward drift (e.g., Engle 1994, Engle and Hamilton 1990). As a consequence, there will tend to be runs in one direction and then the other in the changes of the exchange rate. Such persistence pattern in the direction of changes is thus predictable. From an econometric perspective, the direction of asset returns is predictable using past returns if the conditional mean of asset returns is time-varying (i.e., when the market is not efficient). Christoffersen and Diebold (2002) show that even if the conditional mean is not predictable (i.e., the market is efficient), directional predictability can be driven solely from volatility clustering, as long as the long-run average asset return is nonzero. Breen, Glosten and Jagannathan (1989, p.1184) also point out that given a positive expected excess return, the probability of an up market is a function of both conditional mean and conditional variance. Some empirical works, based on various models and technical trading rules, appear to suggest that it is easier to forecast the direction of asset returns than the level of asset returns (e.g., Breen, Glosten and Jagannathan 1989,

[1] There has been no unified definition of market contagion (see, Stulz 2001). Here, we use the definition that the link between two markets becomes stronger when contagion occurs.

Engle 1994, Kuan and Liu 1995, Larsen and Wozniak 1995, Leitch and Tanner 1991, 1995, Pesaran and Timmermann 1995, 2000, Satchell 1995).

The purpose of this paper is three-folds. First, we propose a model-free omnibus statistical test and a class of separate inference procedures to check whether the direction of asset returns is predictable using currently available information, and if so, what are possible sources of directional predictability. The proposed procedures are based on a generalized cross-spectrum, which extends Hong's (1999) univariate generalized spectrum. The generalized spectrum is the synthesis of the characteristic function and spectral analysis. Because of the use of the characteristic function, the generalized spectrum can capture both linear and nonlinear serial dependence in the data. This is particularly suitable for testing directional predictability because the probability of the direction of changes in an underlying variable generally depends on the dynamics in every conditional moment and is a highly nonlinear function of the history of past changes. At the meantime, the generalized spectrum maintains the nice feature of conventional power spectrum. It can check many lags simultaneously. This is very useful when directional dependence is small at each lag but carries over a very long distributional lag. The omnibus directional predictability test can detect a wide range of alternatives, while the separate inference procedures can check whether the direction of changes can be predicted using the level, volatility, skewness, and kurtosis of past asset returns.

Second, we apply the proposed procedures to a variety of daily U.S. stock price indices—Dow Jones Industrial Averages (DJIA), S&P 500, NASDAQ, and NYSE composite index. We find overwhelming evidence on directional predictability for the excess stock returns. We then explore possible sources of the documented directional predictability of excess stock returns. It is found that the levels of past returns or their directions can be used to predict the direction of future returns with any threshold (including zero). In addition, past volatility clustering can be used to the predict direction of large returns, although not for returns with zero threshold. The documented directional predictability cannot be completely explained by an MA(1)-threshold GARCH(1,1) model.

Third, to show whether the documented directional predictability is useful in practice, we consider a class of autologit models that forecast the 1-step-ahead direction in stock price changes using the direction, level, volatility, skewness, and kurtosis of past price changes respectively. We find that trading strategies based on the 1-step-ahead combined directional forecasts of these autologit models have significantly higher Sharpe ratios than the buy-and-hold trading strategy.

The plan of the paper is organized as follows. In Section 2 we describe the hypotheses of interest and discuss the relationship between directional predictability and the efficient market hypothesis, volatility clustering, as well as serial dependence in higher order conditional

moments such as skewness and kurtosis. Because directional predictability depends on the conditional probability of asset returns exceeding a threshold, serial dependence in variance and higher order moments may also lead to directional predictability even when the market is efficient. In Section 3, we propose a generalized cross-spectral approach to develop a model-free omnibus test for directional predictability and a variety of generalized cross-spectral derivative tests to gauge possible sources of directional predictability. To assess the reliability of the asymptotic distribution theory in finite samples, Section 4 presents a limited simulation study on the finite sample performance of the proposed tests. In Section 5 we apply the tests to a variety of daily stock price indices. In Section 6, we investigate out-of-sample directional predictability of a class of autologit models and their economic significance in terms of extra risk-adjusted trading profit over the buy-and-hold strategy. Section 7 concludes. All mathematical proofs are collected in an appendix.

2 Hypotheses of Interest

Suppose $\{Y_t\}$ is a strictly stationary time series such as a sequence of asset returns. We are interested in whether the directions of future asset returns are predictable using current and past returns. Define the direction indicator function

$$Z_t(c) = \mathbf{1}(Y_t > c), \qquad -\infty < c < \infty,$$

where $\mathbf{1}(\cdot)$ is the indicator function, and c is a threshold constant. Without loss of generality, we can define c in terms of the multiples of the standard deviation $\sigma_Y = \sqrt{\text{var}(Y_t)}$. When $c = 0$, $Z_t(c)$ is an indicator for positive returns. When $c = 1$ (say), $Z(c)$ is an indicator for "large" positive returns. Similarly we can define the directions for negative returns and large negative returns respectively. The later are useful in characterizing large downside risk (e.g., Ang and Chen 2002). The serial dependence structures for small and large returns may be different (Sonik 2001). It is sometime believed that the strength of serial dependence between large returns is stronger than that between small returns, as is the case of market contagion. Investors may be more interested in directional predictability of large asset returns. They may perceive large shocks as containing significant informational contents and small shocks as mere background noises. Consequently, their valuations and expectations react only to large shocks. Moreover, the fact that the number of incorrect forecasts exceeds that of correct forecasts is not necessary to rule out profitability of a trading strategy. A profitable trading strategy may be marked by a small number of successful forecasts for which large profits are made, and a large number of incorrect forecasts for which small losses are incurred (e.g., Cumby and Modest 1987, Diebold and Lopez 1996). Some technical trading rules such as filters do involve

prediction of the direction of returns with certain threshold and their profitability depends on the magnitude of the actual changes.

Let $I_{t-1} \equiv \{Y_{t-1}, Y_{t-2}, ...\}$ be the information set of asset returns available at time $t-1$. The hypotheses of interest are

$$\mathbb{H}_0 : \Pr\{E[Z_t(c)|I_{t-1}] = E[Z_t(c)]\} = 1$$

versus

$$\mathbb{H}_A : \Pr\{E[Z_t(c)|I_{t-1}] = E[Z_t(c)]\} < 1.$$

Note that $E[Z_t(c)|I_{t-1}] = P(Y_t > c|I_{t-1})$ and $E[Z_t(c)] = P(Y_t > c)$. Under $\mathbb{H}_0$, the information set I_{t-1} is useless in predicting the direction of returns with threshold c. In other words, past returns cannot be used to predict the direction of future returns. Under $\mathbb{H}_A$, the direction of returns with threshold c is predictable using the information set I_{t-1}. Note that it is important to specify the threshold constant c because it is possible to predict the direction of returns with some threshold but not with another threshold; see an example below.

The null hypothesis $\mathbb{H}_0$ differs from the efficient market hypothesis;[2] the latter is defined as

$$E(Y_t|I_{t-1}) = \mu \text{ almost surely (a.s.) for some constant } \mu \in (-\infty, \infty).$$

When the market is efficient, the level or the conditional mean of future returns is not predictable using past returns. No systematic trading strategy can be more profitable in the long-run than holding the market portfolio, though of course one can still temporarily beat the market through sheer luck. Market efficiency however, does not necessarily imply that the direction of returns is not predictable. Christoferssen and Diebold (2002) have an excellent theoretic discussion on the relationships among market efficiency, directional predictability, and volatility clustering in a framework where the threshold $c = 0$ and the unconditional mean $\mu \neq 0$. They focus on directional predictability under market efficiency. We now provide some related discussion in our framework.

2.1: Directional Predictability when the Market is Inefficient

When the market is not efficient, the conditional mean $E(Y_t|I_{t-1})$ is a function of I_{t-1} and the level of returns is thus predictable using past returns. In this case, it is generally possible to predict the direction of returns. To see this, consider the following data generating process

$$Y_t = \mu_t + \sqrt{h_t}\varepsilon_t, \tag{2.1}$$

[2] Like Christoffersen and Diebold (2002) as well as the majority of the financial literature (e.g., Fama 1970, Campbell, Lo and MacKinlay 1997), our definition of market efficiency differs from general equlibrium definitions of market efficiency. The latter may be consistent with a predictable time-varying conditional mean due to the presence of time-varying risk premium (e.g., Lucas 1978).

where $\mu_t = E(Y_t|I_{t-1}), h_t = \text{var}(Y_t|I_{t-1})$ and the innovation $\{\varepsilon_t\}$ is an martingale difference sequence with mean 0, variance 1, and conditional CDF $F_\varepsilon(\cdot|I_{t-1})$. Note that there may exist serial dependence in third order or other higher order moments of $\{\varepsilon_t\}$. This is called the weak form volatility process in the literature.[3] Example are Hansen's (1994) autoregressive conditional density model and Harvey and Siddque's (2000) conditional skewness model where ε_t follows an asymmetric Student's t-distribution with time-varying degrees of freedom and skewness. The functions μ_t and h_t characterize serial dependence in the first two conditional moments respectively, while $F_\varepsilon(\cdot|I_{t-1})$ characterizes serial dependence in higher order conditional moments. As will be seen shortly, serial dependence in any moment may affect directional predictability.

Under (2.1), the direction indicator $Z_t(c) = \mathbf{1}[\varepsilon_t > (c-\mu_t)/\sqrt{h_t}]$. Thus, we have

$$E\left[Z_t(c)|I_{t-1}\right] = 1 - F_y\left(c|I_{t-1}\right) = 1 - F_\varepsilon\left(\frac{c-\mu_t}{\sqrt{h_t}}|I_{t-1}\right),$$

where $F_y(\cdot|I_{t-1})$ is the conditional CDF of Y_t given I_{t-1}. As long as μ_t is time-varying and $(c-\mu_t)/\sqrt{h_t}$ is not constant for all t,[4] $E[Z_t(c)|I_{t-1}]$ is a time-varying function no matter whether the threshold $c = 0, h_t$ is a constant, or $\{\varepsilon_t\}$ is i.i.d. Thus, the direction of returns with any threshold c is predictable when μ_t is time-varying (i.e., when the market is not efficient.) Many technical trading rules proposed in the literature, such as those based on artificial neural network models, are based on the directional predictive ability of a conditional mean model.

2.2: Directional Predictability Under Market Efficiency

We now investigate the relationship between the market efficiency and directional predictability. This is of practical importance because it is well-known that there exists little or weak serial dependence in the conditional mean of high-frequency (e.g., daily) financial returns. When the market is efficient ($\mu_t = \mu$ for all t), the direction of returns may or may not be predictable using the information set I_{t-1}, and both cases may be not inconsistent with the efficient market hypothesis. Based on a Gram-Charlier expansion, Christoffersen and Diebold (2002) show that directional dependence does not imply market inefficiency; directional dependence can occur through the interaction between a nonzero unconditional mean μ and volatility clustering.

2.2.1: Threshold c differs from the long run average return μ

First, we consider directional predictability with $c \neq \mu$, i.e., the directional predictability of returns with threshold c different from the long-run average return μ. This is the case

[3] Drost and Nijman (1993) call a GARCH with an i.i.d. noonvation sequence $\{\varepsilon_t\}$ a "strong form GARCH," and a GARCH with non-i.i.d. innovations a "weak form GARCH".

[4] The possibility that $(c-\mu_t)/h_t^{1/2}$ is constant for all t may arise when $\mu_t = c + \alpha h_t^{1/2}$ for some constant α, a specific ARCH-in-Mean process. For ARCH-in-mean model, see Engle, Lilian and Ng (1987).

thoroughly examined in Christoferssen and Diebold (2002) where they assume $c = 0$ and $\mu \neq 0$. As long as the gap $\delta \equiv c - \mu \neq 0$, the directional predictability $E[Z_t(c)|I_{t-1}]$ depends on I_{t-1} via volatility clustering:

$$E\left[Z_t(c)|I_{t-1}\right] = 1 - F_\varepsilon\left(\frac{\delta}{\sqrt{h_t}}|I_{t-1}\right).$$

Thus, $Z_t(c)$ is predictable using the information set I_{t-1}, even if the innovation $\{\varepsilon_t\}$ is i.i.d. so that the conditional CDF $F_\varepsilon(\cdot|I_{t-1}) = F_\varepsilon(\cdot)$ does not depend on I_{t-1}. In this case, the sources of directional predictability solely comes from volatility clustering. Of course, directional predictability can also arise from third order or higher conditional moments of $\{\varepsilon_t\}$, when $\{\varepsilon_t\}$ is not i.i.d. The fact that directional predictability comes from the conditional variance and other higher order conditional moments may explain why it is easier to predict the direction than the level of the change itself, as many empirical studies usually conclude.

2.2.2: Threshold $c = \mu$

An interesting case arises when the gap $\delta \equiv c - \mu = 0$. Here, we have

$$E\left[Z_t(c)|I_{t-1}\right] = 1 - F_\varepsilon\left(0|I_{t-1}\right).$$

Suppose ε_t is i.i.d. so that $F_\varepsilon(\cdot|I_{t-1}) = F_\varepsilon(\cdot)$ is not time-varying. Under market efficiency, serial dependence of returns $\{Y_t\}$ is completely characterized by its conditional variance. In this case, $E\left[Z_t(c)|I_{t-1}\right]$ is not predictable, because $E[Z_t(c)|I_{t-1}] = 1 - F_\varepsilon(0)$ is constant for all t. This is quite different from the case with $\delta \neq 0$, where volatility clustering alone can lead to directional predictability via its interaction with a nonzero δ. In fact, even if $\{\varepsilon_t\}$ is not i.i.d. but has a conditional symmetric distribution (i.e., $F_\varepsilon(-\varepsilon|I_{t-1}) = F_\varepsilon(\varepsilon|I_{t-1})$ for all ε), the direction of return Y_t with threshold $c = \mu$ is not predictable using I_{t-1}.

Next, suppose the gap $\delta = 0$ but $F_\varepsilon(\cdot|I_{t-1})$ is time-varying and is not asymmetric about zero. This suggests that there exists serial dependence in third order and/or higher order conditional moments of $\{\varepsilon_t\}$. In this case, the direction of $Z_t(c)$ is predictable using I_{t-1} and the source of predictability comes from higher order dependence rather than volatility clustering (e.g., conditional skewness and kurtosis). Hansen (1994) and Harvey and Siddque (1999, 2000) find that the conditional skewness of asset returns is time-varying and therefore predictable. This can be another deriving force for the directional predictability of asset returns.

To sum up, (i) when the market is not efficient (i.e., there exists serial dependence in conditional mean), the direction of returns with any threshold c is generally predictable using past returns. (ii) When the market is efficient but there exists serial dependence in such higher order conditional moments as skewness and kurtosis, the direction of returns with any threshold c is also predictable using I_{t-1}. (iii) When the market is efficient and serial dependence is completely characterized by volatility clustering, the direction of return Y_t is predictable using

I_{t-1} except for threshold $c = \mu$. As long as $c \neq \mu$, volatility clustering is a driving force for directional predictability.

3 Tests for Directional Predictability

The above analysis shows that the dynamics of directional predictability of asset returns is highly nonlinear, because it essentially depends on all time-varying conditional moments. We now extend Hong's (1999) generalized spectrum to construct a model-free test for directional predictability. The generalized spectrum of Hong (1999) is particularly suitable for nonlinear time series analysis, thanks to the use of the characteristic function.

Once predictability of the direction of asset returns is documented, it will be interesting and important to gauge possible sources of directional predictability. In particular, one may like to ask whether the direction, level, volatility, skewness, and kurtosis can be used to predict the direction of asset returns. This will provide very useful information for modelling and forecasting the direction of returns. The generalized spectrum can be differentiated to yield such separate inference procedures. This is made possible because the characteristic function can be differentiated to give various moments. We now discuss this econometric methodology.

3.1 Generalized Spectrum

To capture generic serial dependence of a strictly stationary process $\{Y_t\}$ and to explore the pattern of serial dependence of $\{Y_t\}$, Hong (1999), in an univariate time series context, proposes a generalized spectrum as an analytic tool for linear and nonlinear time series. The basic idea is to transform $\{Y_t\}$ via a complex-valued exponential function

$$Y_t \to \exp(\mathbf{i}uY_t), \quad u \in (-\infty, \infty), \mathbf{i} = \sqrt{-1},$$

and then consider the spectrum of the transformed series. Let

$$\varphi(u) \equiv E\left(e^{\mathbf{i}uY_t}\right)$$

be the marginal characteristic function of $\{Y_t\}$ and let

$$\varphi_j(u, v) \equiv E\left[e^{\mathbf{i}(uY_t + vY_{t-|j|})}\right], \quad j = 0, \pm 1, ...$$

be the pairwise joint characteristic function of $(Y_t, Y_{t-|j|})$, where j is a lag/lead order. Define the covariance function between the transformed variables $e^{\mathbf{i}uY_t}$ and $e^{\mathbf{i}vY_{t-|j|}}$:

$$\sigma_j(u, v) \equiv \text{cov}(e^{\mathbf{i}uY_t}, e^{\mathbf{i}vY_{t-|j|}}), \quad u, v \in (-\infty, \infty). \tag{3.1}$$

Straightforward algebra yields $\sigma_j(u,v) = \varphi_j(u,v) - \varphi(u)\varphi(v)$. Because $\varphi_j(u,v) = \varphi(u)\varphi(v)$ for all $u, v \in (-\infty, \infty)$ if and only if Y_t and $Y_{t-|j|}$ are independent, $\sigma_j(u,v)$ can capture any type of pairwise serial dependence over various lags in $\{Y_t\}$, including those with zero autocorrelation. It is well-known that many high-frequency financial time series display little serial correlation but persistent volatility clustering and other higher order dependence.

Under suitable conditions, the Fourier transform of $\sigma_j(u,v)$ exists and is given by:

$$f_{ZZ}(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v) e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi, \pi], \tag{3.2}$$

where ω is frequency. Like $\sigma_j(u,v)$, $f_{ZZ}(\omega,u,v)$ can capture any type of pairwise serial dependencies in $\{Y_t\}$ over various lags (i.e., dependence between Y_t and Y_{t-j} for any $j \neq 0$). Unlike the power spectrum and higher order spectra (e.g., bispectrum),[5] the generalized spectrum $f_{ZZ}(\omega,u,v)$ does not require any moment condition on $\{Y_t\}$. In other words, $\{Y_t\}$ may not be weakly stationary (e.g., when $\{Y_t\}$ is an integrated GARCH process; see Bollerslev and Engle 1986). This is appealing in finance because it is often argued (e.g., Pagan and Schwert 1990) that certain moments like the unconditional variance of some high frequency financial time series may not exist.

When $\text{var}(Y_t) = \sigma_Y^2$ exists, the conventional power spectrum of $\{Y_t\}$ can be obtained by differentiating $f_{ZZ}(\omega,u,v)$ with respect to (u,v) at $(0,0)$:

$$h_{YY}(\omega) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \text{cov}(Y_t, Y_{t-|j|}) e^{-\mathbf{i}j\omega} = -\frac{\partial^2}{\partial u \partial v} f_{YY}(\omega,u,v) \,\big|_{(u,v)=(0,0)} .$$

For this reason, $f_{YY}(\omega,u,v)$ is called a "generalized spectral density" of $\{Y_t\}$.

When all the moments of $\{Y_t\}$ exist, we can decompose, by a Taylor series expansion, the generalized spectrum as follows:

$$f_{YY}(\omega,u,v) = \sum_{m=1}^{\infty} \sum_{l=1}^{\infty} \frac{\mathbf{i}^{m+l}}{m!l!} \left[\frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \text{cov}(Y_t^m, Y_{t-|j|}^l) e^{-\mathbf{i}j\omega} \right] u^m v^l,$$

where (u,v) is near $(0,0)$. This indicates that $f_{YY}(\omega,u,v)$ can captures various correlations between Y_t^m and $Y_{t-|j|}^l$ for all $m, l > 0$. Of course, $f_{YY}(\omega,u,v)$ does not require existence of moments of $\{Y_t\}$.

The introduction of transform parameters (u,v) provides much flexibility for $f_{YY}(\omega,u,v)$ to capture serial dependence in $\{Y_t\}$. For example, the supremum generalized spectrum $s_{YY}(\omega) \equiv \sup_{-\infty<u,v<\infty} |f_{YY}(\omega,u,v)|$ can be viewed as the maximum serial dependence of

[5] For conventional power spectral analysis, see (e.g.) Priestley (1981). For bispectral analysis, see (e.g.) Sub Rao and Gabr (1984).

$\{Y_t\}$ at frequency ω. This can be used to identify business cycles, seasonalities (e.g., calender effects), or other forms of periodicities caused by linear or nonlinear dependence (e.g., persistent volatility clustering).

The generalized spectrum is a synthesis of spectral analysis and the characteristic function. Spectral analysis is not uncommon in economics and finance (e.g., Durlauf 1990, Granger 1969, Watson 1993). An advantage of spectral analysis is that it includes information of all lags simultaneously in a natural manner. On the other hand, the characteristic function can capture linear and nonlinear dependencies (including those with zero autocorrelation), thus overcoming the drawback of the conventional power spectrum. As a consequence, generalized spectrum is particularly suitable for analyzing complex and nonlinear economic and financial systems. We note that there has been an increasing interest in using the characteristic function in economics and finance. Among them are Hong and Lee (2003), Jiang and Knight (2002), Knight and Yu (2002), Pinkse (1998), and Singleton (2001).

3.2 Generalized Cross-Spectrum

The generalized spectrum $f_{YY}(\omega, u, v)$ of $\{Y_t\}$ is useful in exploring how Y_t depends on its own past history I_{t-1}. It cannot, however, be directly applied to investigate whether the direction of returns, $Z_t(c) \equiv \mathbf{1}(Y_t > c)$, is predictable using the information set I_{t-1}. For this purpose, we have to expend Hong's (1999) univariate generalized spectral analysis to a bivariate generalized cross-spectral analysis. Suppose $\{Z_t, Y_t\}$ is a bivariate strictly stationary time series process with marginal characteristic functions $\phi_{_Z}(u) = E(e^{iuZ_t})$ and $\phi_{_Y}(u) = E(e^{iuY_t})$. In this subsection, we permit but do not require $Z_t = Z_t(c)$. Define the generalized cross-covariance function

$$\sigma_{_{ZY},j}(u, v) \equiv \text{cov}\left(e^{\mathbf{i}uZ_t}, e^{\mathbf{i}vY_{t-|j|}}\right), \quad j = 0, \pm 1, \ldots . \tag{3.3}$$

Straightforward algebra shows

$$\sigma_{_{ZY},j}(u, v) = \phi_{_{ZY},j}(u, v) - \phi_{_Z}(u)\phi_{_Y}(v),$$

where $\phi_{_{ZY},j}(u, v) \equiv E[e^{i(uZ_t + vY_{t-|j|})}]$ is the joint characteristic function of $(Z_t, Y_{t-|j|})$. Because $\sigma_{_{ZY},j}(u, v) = 0$ for all $u, v \in (-\infty, \infty)$ if and only if Z_t and $Y_{t-|j|}$ are independent, $\sigma_{_{ZY},j}(u, v)$ can capture any type of cross-dependence between Z_t and $Y_{t-|j|}$.

Analogous to the univariate generalized spectrum $f_{YY}(\omega, u, v)$, we may call the Fourier transform of $\sigma_{_{ZY},j}(u, v)$,

$$f_{_{ZY}}(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_{_{ZY},j}(u, v) e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi, \pi], \tag{3.4}$$

the "generalized cross-spectral density" between $\{Z_t\}$ and $\{Y_{t-j}, j > 0\}$. Like $\sigma_{ZY,j}(u,v)$, $f_{ZY}(\omega,u,v)$ can capture any type of pairwise cross-dependence between Z_t and $Y_{t-|j|}$. It can be used to explore how Z_t depends on the entire past history of $\{Y_t\}$. In particular, it can be used to examine various linear and nonlinear Granger causalities from lagged variables $\{Y_{t-|j|}\}$ to Z_t.[6] Note that no moment condition on $\{Y_t\}$ and $\{Z_t\}$ is needed for $f_{ZY}(\omega,u,v)$.

We now consider a special case relevant to our interest of directional predictability. Suppose Z_t is independent of I_{t-1}. Then the generalized cross-spectrum $f_{ZY}(\omega,u,v)$ becomes a flat generalized cross-spectrum:

$$f_{ZY,0}(\omega,u,v) \equiv \frac{1}{2\pi}\sigma_{ZY,0}(u,v), \quad \omega \in [-\pi,\pi]. \tag{3.5}$$

Thus, one can test independence between Z_t and $\{Y_{t-j}, j>0\}$ by comparing $f_{ZY}(\omega,u,v)$ and $f_{ZY,0}(\omega,u,v)$. Any significant difference between them will indicate the dependence of Z_t on the past history of $\{Y_t\}$.

Just as the characteristic function can be differentiated to generate various moments (when they exist), $f_{ZY}(\omega,u,v)$ can be differentiated to capture various cross-dependencies between Z_t and $\{Y_{t-j}, j>0\}$. Consider the following generalized cross-spectral density derivative

$$f_{ZY}^{(0,m,l)}(\omega,u,v) \equiv \frac{\partial^{m+l}}{\partial u^m \partial v^l} f_{ZY}(\omega,u,v) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \sigma_{ZY,j}^{(m,l)}(u,v)e^{-\mathbf{i}j\omega}, \quad m,l \geq 0. \tag{3.6}$$

Such a derivative exists provided $E|Z_t|^{2m} < \infty$ and $E|Y_t|^{2l} < \infty$. To check $E(Z_t|I_{t-1}) = E(Z_t)$ a.s., as is the hypothesis of interest $\mathbb{H}_0$ in testing directional predictability (with $Z_t = Z_t(c)$), we can use the $(1,0)$-th order generalized cross-spectral derivative

$$f_{ZY}^{(0,1,0)}(\omega,0,v) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \sigma_{ZY,j}^{(1,0)}(0,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi], \tag{3.7}$$

where $\sigma_{ZY,j}^{(1,0)}(0,v) = \text{cov}(\mathbf{i}Z_t, e^{\mathbf{i}vY_{t-|j|}})$. This $(1,0)$-derivative essentially checks whether

$$E(Z_t|Y_{t-|j|}) = E(Z_t), \quad j = 0, \pm 1, ...$$

because $\sigma_{zy,j}^{(1,0)}(0,v) = 0$ if and only if $E(Z_t|Y_{t-|j|}) = E(Z_t)$ a.s. under suitable conditions. The latter is similar in spirit to the null hypothesis $\mathbb{H}_0$ when $Z_t = Z_t(c)$.[7] Intuitively, $\sigma_{ZY,j}^{(1,0)}(0,v)$ can capture correlations between Z_t and all moments of $Y_{t-|j|}$, thus exploiting all implications of $E(Z_t|Y_{t-|j|}) = E(Z_t)$. Therefore, with $Z_t = Z_t(c)$, we can use $f_{ZY}^{(0,1,0)}(\omega,0,v)$ to check the directional predictability hypotheses $\mathbb{H}_0$ versus $\mathbb{H}_A$.

[6] For general Granger causality, see Granger (1980).

[7] See Bierens (1982) and Stinchcombe and White (1998) for related discussion and proof in a different context. Bierens (1982) and Stinchcombe and White (1998) consider specification for regression models where Z_t is the regression model error and Y_t is the regressor vector.

Once directional predictability is documented using $f_{ZY}^{(0,1,0)}(\omega,0,v)$, one may like to further explore possible sources for directional predictability. In particular, is directional predictability caused by conditional mean dynamics? Or is it caused by volatility clustering? Or is it caused by conditional skewness or other higher order conditional moment? Such information will be very helpful for making inferences on the nature of directional predictability and providing useful guidance in constructing directional forecast models.

To gauge possible reasons of directional predictability, we can use higher order generalized cross-spectral derivative

$$f_{ZY}^{(0,1,l)}(\omega,0,0)=\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_{ZY,j}^{(1,l)}(0,0)e^{-\mathbf{i}j\omega},\qquad \omega\in[-\pi,\pi], \tag{3.8}$$

where $\sigma_{ZY,j}^{(1,l)}(0,0)=\text{cov}\left[\mathbf{i}Z_t,(\mathbf{i}Y_{t-|j|})^l\right], l\geq 1$. For $l=1,2,3,4, \sigma_{ZY}^{(1,l)}(0,0)$ will be proportional to cross-covariances $\text{cov}(Z_t,Y_{t-|j|}^l)$. As a consequence, we can use $f_{ZY}^{(0,1,l)}(\omega,0,0)$ to check whether Z_t is predictable using the level of past changes $\{Y_{t-j}\}$, past volatility $\{Y_{t-j}^2\}$, past skewness $\{Y_{t-j}^3\}$ and past kurtosis $\{Y_{t-j}^4\}$ respectively. Below, we will develop a unified framework that includes all of these tests using various generalized cross-spectral derivatives.

3.3 Generalized Cross-Spectral Tests for Directional Predictability

Suppose we have a random sample of asset returns $\{Y_t\}_{t=1}^T$ of size T. Define the empirical generalized cross-covariance function between $\{Z_t(c)\}$ and $\{Y_t\}$

$$\hat{\sigma}_{ZY,j}(u,v)=\hat{\phi}_{ZY}(j,u,v)-\hat{\phi}_{ZY}(j,u,0)\hat{\phi}_{ZY}(j,0,v), \tag{3.9}$$

where

$$\hat{\phi}_{ZY}(j,u,v)=(T-|j|)^{-1}\sum_{t=|j|+1}^{T}e^{\mathbf{i}(uZ_t(c)+vY_{t-|j|})},\qquad j=0,\pm1,...,\pm(T-1),$$

is the empirical joint characteristic function for $\{Z_t(c),Y_{t-|j|}\}$. To estimate the generalized cross-spectral density $f_{ZY}(\omega,u,v)$ in (3.4), we use a smoothed kernel estimator:

$$\hat{f}_{ZY}(\omega,u,v)=\frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{1/2}k(j/p)\hat{\sigma}_{ZY,j}(u,v)e^{-\mathbf{i}j\omega}. \tag{3.10}$$

Here, $k(\cdot)$ is a kernel function that assigns weights to various lags. It can have bounded support. An example is the Bartlett kernel $k(z)=(1-|z|)\,\mathbf{1}\,(|z|\leq 1)$, which is popular in econometrics (cf. Newey and West 1987). In this case, p is the maximum lag truncation order. The kernel $k(\cdot)$ can also have unbounded support. An example is the Daniell kernel $k(z)=\sin(\pi z)/(\pi z), -\infty<z<\infty$. In this case, p is no longer a lag truncation number but a

smoothing parameter that governs the smoothness of the spectral estimator $\hat{f}_{ZY}(\omega, u, v)$. We can view that p is an effective lag order because lags much larger than p receive little weights. The factor $(1 - |j|/T)^{1/2}$ in (3.10) is a finite sample correction factor. It could be replaced by unity without affecting consistent estimation of the generalized cross-spectrum $f_{ZY}(\omega, u, v)$, but it gives better finite sample performance for the proposed tests below. Under proper conditions on the kernel $k(\cdot)$ and the lag order p, as well as on serial dependence of $\{Y_t\}$, it can be shown that the estimator $f_{ZY}(\omega, u, v)$ is consistent for $f_{ZY}(\omega, u, v)$.[8] The generalized spectral approach has at least three appealing features: First, $\hat{f}_{ZY}(\omega, u, v)$ employs many lags simultaneously because it is usually required that $p \equiv p(T) \to \infty$ as $T \to \infty$. In particular, when $k(\cdot)$ has infinite support, all $T-1$ lags available in the sample are used. This is expected to have good power in detecting cross-dependence that decays to zero slowly as the lag order j increases. Second, the kernel function $k(\cdot)$ provides a natural weighting scheme for various lags. Typically, higher order lags are discounted, which may enhance the power of the proposed tests in practice because financial markets are more influenced by the recent events than by the remote events remote past events. Third, one can choose a lag order p via suitable data-driven methods. For example, we can select a data-driven p that minimizes the integrated mean squared error of the generalized cross-spectral density estimator $\hat{f}_{ZY}(\omega, u, v)$. See Hong (1999) for more discussion in the context of univariate generalized spectrum.

To check directional predictability and its possible sources, we shall compare the generalized cross-spectral derivative estimators

$$\hat{f}_{ZY}^{(0,1,l)}(\omega, 0, v) = \frac{\partial^{1+l}}{\partial u \partial v^l} \hat{f}_{ZY}(\omega, 0, v) \tag{3.11}$$

and

$$\hat{f}_{ZY,0}^{(0,1,l)}(\omega, 0, v) = \frac{\partial^{1+l}}{\partial u \partial v^l} \hat{f}_{ZY,0}(\omega, 0, v), \tag{3.12}$$

where $\hat{f}_{ZY,0}(\omega, u, v) \equiv (2\pi)^{-1} \hat{\sigma}_{ZY,0}(u, v)$ is a consistent estimator for $f_{ZY,0}(\omega, u, v)$; the latter is a flat cross-spectrum implied by the null hypothesis $\mathbb{H}_0$ of no directional predictability. A significant difference between $\hat{f}_{ZY}^{(0,1,l)}(\omega, 0, v)$ and $\hat{f}_{ZY,0}^{(0,1,l)}(\omega, 0, v)$ will indicate directional predictability. To measure the discrepancy between $\hat{f}_{ZY}^{(0,1,l)}(\omega, 0, v)$ and $\hat{f}_{ZY,0}^{(0,1,l)}(\omega, 0, v)$, we can use a convenient quadratic form

$$\begin{aligned} \hat{Q}(1, l) &= \pi T \int \int_{-\pi}^{\pi} |\hat{f}_{ZY}^{(0,1,l)}(\omega, 0, v) - \hat{f}_{ZY,0}^{(0,1,l)}(\omega, 0, v)|^2 d\omega dW(v) \\ &= \sum_{j=1}^{T-1} k^2(j/p)(T - j) \int |\hat{\sigma}_{ZY,j}^{(1,l)}(0, v)|^2 dW(v), \end{aligned} \tag{3.13}$$

[8] See Hong (1999, Theorem 1). Although the present context is a bivariate framework while Hong (1999) considers a univariate process; the proof and regularity conditions are similar.

where $W(\cdot)$ is a positive and nondecreasing weighting function with bounded total variation and the unspecified integral is taken over the support of $W(\cdot)$. An example of $W(\cdot)$ is $W(\cdot) = \Phi(\cdot)$, the N(0,1) cdf, commonly used in the empirical characteristic function literature. Note that there is no need to calculate the integration over frequency ω, but we still need to calculate the integration over v. The integral over v can be calculated using numerical integration methods, such as Gauss-Quadrature, available in most statistical software. For accurate numerical integration, we can truncate the N(0,1) CDF on a bounded support, say [-3,3]. There is no requirement that $W(v)$ be integrated to 1.

Our test statistic is a standardized version of the cumulative sum of $\hat{Q}(1,l)$:

$$M_{ZY}(1,l) = \left[\hat{Q}(1,l) - \hat{C}_{ZY}(1,l)\sum_{j=1}^{T-1} k^2(j/p)\right] / \left[\hat{D}_{ZY}(1,l)\right]^{1/2}, \tag{3.14}$$

where the integer $l \geq 0$, the centering and scaling factors

$$\begin{aligned}
\hat{C}_{ZY}(1,l) &= \hat{\lambda}(c)[1-\hat{\lambda}(c)] \int |\hat{\sigma}^{(l,l)}_{YY,0}(v,-v)|^2 dW(v), \\
\hat{D}_{ZY}(1,l) &= 2\hat{\lambda}^2(c)[1-\hat{\lambda}(c)]^2 \sum_{j=1}^{T-2}\sum_{\tau=1}^{T-2} k^2(j/p)k^2(\tau/p) \iint |\hat{\sigma}^{(l,l)}_{YY,j-\tau}(v,v')|^2 dW(v)dW(v'),
\end{aligned}$$

$\hat{\lambda}(c) = T^{-1}\sum_{t=1}^{T} Z_t(c)$ is the sample proportion for $\{Y_t > c\}$, and $\hat{\sigma}_{YY,j}(v,v')$ is the empirical generalized autocovariance function of $\{Y_t\}$; namely,

$$\hat{\sigma}_{YY,j}(v,v') = \hat{\phi}_{YY,j}(v,v') - \hat{\phi}_{YY,j}(v,0)\hat{\phi}_{YY,j}(0,v')$$

and $\hat{\phi}_{YY,j}(v,v') = (T-|j|)^{-1}\sum_{t=|j|+1}^{T} e^{\mathbf{i}(vY_t + v'Y_{t-|j|})}$. Note that the factors $\hat{C}_{ZY}(1,l)$ and $\hat{D}_{ZY}(1,l)$ have taken into account generic serial dependence within $\{Y_t\}$, which is present even when $\mathbb{H}_0$ holds. Intuitively, $\hat{f}_{ZY,0}(\omega,0,v)$ is an efficient estimator for $f_{ZY}(\omega,u,v)$ under $\mathbb{H}_0$, and $\hat{f}_{ZY}(\omega,u,v)$ is an inefficient but consistent estimator for $f_{ZY}(\omega,u,v)$ under $\mathbb{H}_A$. Thus, our test is similar in spirit to Hausman's (1978) test.

Under suitable regularity conditions, we can show that as the lag order $p \equiv p(T) \to \infty, p/T \to 0$, $\hat{M}_{ZY}(1,l)$ converges in distribution to N(0,1) under $\mathbb{H}_0$ and generally diverges to positive infinity under $\mathbb{H}_A$ (see the appendix for the asymptotic theory). Appropriate critical values are the upper-tailed N(0,1) critical values (e.g., 1.65 at the 5% level).

When $l = 0$, $M_{ZY}(1,0)$ is an omnibus test for $\mathbb{H}_0$, because essentially check correlations between $Z_t(c)$ and $Y^l_{t-|j|}$ for all l and j. On the other hand, the separate tests $M_{ZY}(1,l)$ with $l \geq 1$ and $W'(v) = \delta(v)$, the Dirac delta function, can reveal useful information about possible sources for directional predictability.[9] The use of the Dirac delta function implies that we focus

[9] The Dirac delta function $\delta(\cdot)$ is defined as follows: $\delta(u) = 0$ for all $u \neq 0$ and $\int \delta(u)du = 1$.

all weight mass at $v = 0$. As noted earlier, for $l = 1, 2, 3, 4, \sigma_{zy}^{(1,l)}(0,0)$ will be proportional to the cross-covariances $\text{cov}[Z_t(c), Y_{t-|j|}^l]$. Thus, we can use $M_{ZY}(1,l)$ to check whether $Z_t(c)$ is predictable using the level of returns $\{Y_{t-j}\}$, past volatility $\{Y_{t-j}^2\}$, past skewness $\{Y_{t-j}^3\}$ and past kurtosis $\{Y_{t-j}^4\}$ respectively.

On the other hand, one may also like to check if the directions of past returns are helpful in predicting the direction of future returns. This is of interest, for example, when one likes to check if price reversals exist. To test this, we can use the univariate generalized spectral density function of the direction indicator series $\{Z_t(c)\}$,

$$f_{ZZ}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_{ZZ,j}(u,v) e^{-\mathbf{i}j\omega}, \tag{3.15}$$

where the generalized covariance function

$$\sigma_{ZZ,j}(u,v) = \text{cov}\left(e^{\mathbf{i}uZ_t(c)}, e^{\mathbf{i}vZ_{t-|j|}(c)}\right). \tag{3.16}$$

Because $Z_t(c)$ is a Bernoulli random variable taking value 0 or 1, it is straightforward to show that when $Z_t(c)$ is not predictable using I_{t-1}, $Z_t(c)$ is independent of I_{t-1}. One important implication of this is that the sequence of direction indicators, $\{Z_t(c)\}$, is an i.i.d. Bernoulli sequence. Thus, one could test directional predictability by testing i.i.d. for $\{Z_t(c)\}$. If evidence against i.i.d. is found for $\{Z_t(c)\}$, one can conclude that the direction of returns is predictable using the past history of the return directions $\{Z_{t-1}(c), Z_{t-2}(c), ...\}$.

We can test i.i.d. for $\{Z_t(c)\}$ by using the generalized spectral density $f_{ZZ}(\omega, u, v)$ of $\{Z_t(c)\}$. Because $\{Z_t\}$ is an i.i.d. Bernoulli sequence under $\mathbb{H}_0$, the generalized spectrum $f_{ZZ}(\omega, u, v)$ becomes a flat spectrum with respect to frequency ω:

$$f_{ZZ,0}(\omega, u, v) = \frac{1}{2\pi}\sigma_{ZZ,0}(u,v), \quad \omega \in [-\pi, \pi]. \tag{3.17}$$

To test whether the directions of past returns can be used to predict the directions of future returns, we can compare a consistent kernel estimator for $f_{ZZ}(\omega, u, v)$ and a consistent estimator for $f_{ZZ,0}(\omega, u, v)$, defined in the same way as $\hat{f}_{ZY}(\omega, u, v)$ and $\hat{f}_{z_{ZY},0}(\omega, u, v)$ in (3.10) and (3.11) respectively. The associated test is

$$\begin{aligned} M_{ZZ}(0,0) &= \left[\sum_{j=1}^{T-1}(T-j)k^2(j/p)\int |\hat{\sigma}_{ZZ,j}(u,v)|^2 dW(u)dW(v) - \hat{C}_{ZZ}(0,0)\sum_{j=1}^{T-1}k^2(j/p)\right] \\ &\quad \div \left[2\hat{D}_{ZZ}(0,0)\sum_{j=1}^{T-1}k^4(j/p)\right]^{1/2}, \end{aligned} \tag{3.18}$$

where the centering and scaling factors

$$\begin{aligned}\hat{C}_{ZZ}(0,0) &= \left[\int \left[1-|\hat{\phi}_{Z}(v)|^2\right] dW(v)\right]^2, \\ \hat{D}_{ZZ}(0,0) &= \left[\iint \left[\hat{\phi}_{Z}(v+v')-\hat{\phi}_{Z}(v)\hat{\phi}_{Z}(v')\right] dW(v)dW(v')\right]^2,\end{aligned}$$

and $\hat{\phi}_{Z}(v) = T^{-1}\sum_{t=1}^{T} e^{ivZ_t(c)}$ is the empirical characteristic function of $\{Z_t(c)\}$. The test statistic $M_{ZZ}(0,0)$ is a special case covered in Hong (1999). It is asymptotically N(0,1) under $\mathbb{H}_0$. Also, the upper-tailed N(0,1) critical values should be used. A particularly appealing feature of this test is that the validity of the asymptotic distribution of $M_{ZZ}(0,0)$ does not require stationarity of $\{Y_t\}$. Even if $\{Y_t\}$ is not strictly stationary, $\{Z_t\}$ will be still a sequence of i.i.d. Bernoulli random variables under $\mathbb{H}_0$.

An important common feature of the $M_{ZY}(1,l)$ and $M_{ZZ}(0,0)$ tests is that the lag order j is weighted by $k^2(j/p)$. Typically, $k(z)$ gives the largest weight at $z=0$ and smaller weights as $|z|\to\infty$. Thus, higher order lags are discounted. This is expected to enhance power when the current returns are more affected by recent information than by remote information as economic agents digest information available. Another important feature of our spectral approach is that we consider many lags simultaneously by requiring $p\to\infty$ as $T\to\infty$. This is desirable when the dependence of $Z_t(c)$ on $Y_{t-|j|}$ decays to zero slowly as the lag order $j\to\infty$. To implement the test $M_{ZY}(1,l)$ or $\hat{M}_{ZZ}(0,0)$, one has to choose a lag order sequence p. Another advantage of the spectral approach is that the lag order can be chosen via some data-driven methods. Hong (1999) discusses how to choose p via an integrated mean squared error criterion, which trades off between the variance and squared bias of the generalized spectral density estimator. This method still involves the choice of a preliminary "pilot" lag order $\bar{p}$, but the impact of choosing $\bar{p}$ is much smaller. The sampling variation of the data-driven p unavoidably induces additional noises into the test statistics. This adversely affects the size of the tests but it is expected to enhance good power for the tests. We will use it tailored to the present context in both our simulation and empirical applications. Simulation studies show that the performance of the tests are more or less robust to the choice of $\bar{p}$.

4 Finite Sample Performance

The asymptotic N(0,1) distribution of the proposed tests is convenient to use in practice. Before real data applications, however, we need to make sure that it provides reasonable approximations in finite samples. Any inference and conclusion based on a poor asymptotic distribution theory will be misleading about directional predictability of financial time series. For example, suppose a test rejects the correct null hypothesis too often at a given significance

level. Then, when applied to a real data, a significant test statistic would be not reliable because we do not know whether it is due to the poor performance of the test or the true feature of the data.

To assess the finite sample performance, we consider two data generating processes (DGP). DGP1 is an i.i.d.N(0,1) process, and DGP2 is a GARCH(1,1)-i.i.d.N(0,1) process,

$$\begin{cases} Y_t = h_t^{1/2}\varepsilon_t, \\ h_t = 0.05 + 0.8h_{t-1} + 0.15Y_{t-1}^2 \\ \varepsilon_t \sim \text{i.i.d.N(0,1)}, \end{cases}$$

where the GARCH parameter values are the typically empirical estimates for high-frequency financial series (e.g., Bollerslev 1987). Under DGP1, there is no serial dependence in every conditional moment of $\{Y_t\}$. Thus, the direction of returns with any threshold c is not predictable. This allows us to examine the size performance of all the tests for the direction of returns with any threshold c. Under DGP 2, serial dependence exists only in the conditional variance of $\{Y_t\}$. Thus, the direction of returns with threshold $c = 0$ is not predictable using the past returns. However, the directions of returns with nonzero thresholds are predictable under DGP2, due to volatility clustering. Hence, our tests should have nontrivial power whenever c is nonzero.

To compute the statistics $M_{ZY}(1,0)$ and $M_{ZZ}(0,0)$, we use the weighting function $W(\cdot) = \Phi(\cdot)$, the N(0,1) CDF truncated on [-3,3]. We scale both $\{Y_t\}_{t=1}^T$ and $\{Z_t(c)\}_{t=1}^T$ respectively so that they have a unit sample standard deviation. We also use the Bartlett kernel for $k(\cdot)$. To choose a lag order p, we use a procedure analogous to Hong's (1999) plug-in method that is based on the integrated mean squared error criterion of the generalized spectral density estimator. This method also involves the choice of a kernel function and a preliminary lag order $\bar{p}$. We use the Bartlett kernel again. To examine the impact of the choice of preliminary lag order $\bar{p}$, we choose $\bar{p}$ from 11 to 61. This covers a rather wide range of lag orders.

Figures 1 and 2 reports the empirical rejection rates, as a function of $\bar{p}$, of the tests for the direction indicators $Z_t(0) = \mathbf{1}(Y_t > 0), \mathbf{1}(Y_t > 1)$ and $\mathbf{1}(Y_t > 1) - \mathbf{1}(Y_t < -1)$ respectively, under DGP1. Two significance levels, 10% and 5%, with two sample sizes $T = 500, 1,000$, are considered. Overall, the proposed tests perform reasonably well at both the 10% and 5% levels. There are some (but not excessive) overrejections at the 5% level, particularly for $Z_t(0) = \mathbf{1}(Y_t > 0)$ and $\mathbf{1}(Y_t > 1)$. The tests with $Z_t = \mathbf{1}(Y_t > 1) - \mathbf{1}(Y_t < -1)$ have slightly better performance in many scenarios. In general, the sizes of the tests are robust to the choice of $\bar{p}$.

Figures 3 and 4 report the empirical rejection rates of the tests under DGP2 with $T = 500, 1,000$ respectively. First, we consider $c = 0$. Under DGP2, there exists no directional

predictability when and only when $c = 0$. The rejection rates of all the tests are close to the nominal levels of 10% and 5% respectively in this case. We observe that there are more overrejections than under DGP1, but these overrejections are not excessive, particularly in views of our nonparametric time series testing approach with a data-driven lag order selection, which induces additional noise into test statistics. For nonzero thresholds c, all the tests are expected to have power under DGP2 for sufficiently large sample size T because $Z_t(c)$ has directional predictability via the interaction between time-varying volatility and nonzero threshold c. This is indeed the case as shown in Figures 3 and 4. Note that both size and power are relatively robust to the choice of the preliminary lag order $\bar{p}$.

Overall, the simulation evidence shows that the proposed tests have reasonable sizes in finite sample sizes, and have good power against directional predictability.

5 Directional Predictability of Stock Returns

5.1 Data

We now apply our generalized cross-spectral tests to examine directional predictability of a variety of U.S. daily stock price indices, which is essential for macroforecasting and market timing. The stock price indices include Dow Jones average index (DJIA), S&P 500 index (S&P500), NASDAQ composite index (NASDAQ), and NYSE composite index (NYSE). We mainly focus on the directional predictability of excess stock return series:

$$Y_t = 100 \ln(P_t/P_{t-1}) - r_t,$$

where P_t is the daily closing stock price, and r_t, a re-scaled risk-free daily interest rate, is the 3-month treasury bill rate divided by 252, the average trading days in a year. All the stock data are obtained from *Datastream*, and the 3-month T-bill rates are downloaded from the website $www.fed.org$. Table 1 summarizes some basic statistics for the excess returns of all stock indices, which have the same ending date, December 31, 2001, but may have different starting dates for the samples. DJIA and S&P 500 have the largest samples (from February 1, 1962), with 10,043 observations. The sample means of the excess returns of all the indices are nonnegative, but they are very small or close to zero. All excess returns have excessive kurtosis, indicating non-Gaussian features.

5.2 Directional Predictability of Stock Returns

We first check whether the directions are predictable for excess stock returns. Specifically, we consider the predictability of each of the following direction indicators:

$$\begin{aligned} Z_{1t}(c) &= \mathbf{1}(Y_t > c), \\ Z_{2t}(c) &= \mathbf{1}(Y_t < -c), \\ Z_{3t}(c) &= \text{sign}(Y_t, c) = \mathbf{1}(Y_t > c) - \mathbf{1}(Y_t < -c), \end{aligned}$$

for $c = 0, 0.5, 1, 1.5$, in units of sample standard deviation of $\{Y_t\}$. The dynamics of directional predictability can be different between up and down stock markets, and between large and small changes. [10]

Table 2 reports the test statistics $M_{ZY}(1, l)$ for $l = 0, 1, 2, 3, 4$ and $M_{ZZ}(0, 0)$. For the results in all test statistics, we use the Bartlett kernel and a preliminary lag order $\bar{p} = 21$.[11] First, the omnibus test $M_{ZY}(1, 0)$ examines whether the sign (two-sided) direction, positive direction and negative direction respectively, using past excess returns $\{Y_{t-|j|}\}$, are predictable. For sign, positive and negative directions, $M_{ZY}(1, 0)$ is very large for all threshold c (including $c = 0$), essentially implying a zero p-value. There exists overwhelming evidence on the directional predictability for the four indices. For the sign direction, there seems no clear evidence that the direction of large excess returns is easier to predict than the direction of small excess returns. For one-sided (i.e., positive or negative) directions, however, there does exist stronger evidence on the directional predictability of large excess returns than small excess returns (particularly for NASDAQ index), although the $M_{ZY}(1, 0)$ statistic value is not monotonically increasing in threshold c. This suggests that the serial dependence between returns with nonzero thresholds is stronger than the serial dependence between returns with zero threshold. On the other hand, there seems to be a weak evidence that the direction of large positive returns is easier to predict than the direction of large negative returns, using past returns.

We also consider the directional predictability of gross stock returns (i.e., daily price changes $100\ln(P_t/P_{t-1})$ without demeaned by the interest rate r_t); the results (not reported) are very similar to those for the excess returns.

5.3 Sources of Directional Predictability

The finding that the direction of the excess returns of stock indices is predictable using past excess returns is important. However, the omnibus $M_{ZY}(1, 0)$ statistic does not provide any constructive information about possible sources of directional predictability. For this purpose,

[10] Mcqueen, Pinegar and Thorley (1996, p.892), for example, find evidence of different autocorrelation in returns between up and down stock markets.

[11] We also use preliminary lag orders $\bar{p}$ from 11 to 61. The results, which are available from the authors upon request, are similar.

we can use $M_{ZZ}(0,0)$ and the derivative tests $M_{ZY}(1,l)$ for $l = 1,2,3,4$. These tests check whether the direction of future excess returns can be predicted using the direction, level, volatility, skewness and kurtosis of past excess returns.

The $M_{ZY}(1,1)$ statistic examines whether the direction of excess returns can be predicted by the level of past excess returns $Y_{t-|j|}$. For all four indices, the direction of excess returns with any threshold c is predictable using the level of past excess returns. Among other things, the fact that the direction of excess returns with zero threshold can be predicted using the level of past excess returns suggests that a driving force for directional predictability may be a possibly time-varying conditional mean. This is consistent with Lo and MacKinlay (1988) finding that there exists weak serial dependence in the level of stock returns.

There is no clear evidence that large excess returns are easier to predict than smaller ones in direction, using the level of past excess returns. In fact, as threshold c increases, the sign direction of NASDAQ becomes harder to predict when using the level of past excess returns. On the other hand, there exists some evidence that it is easier to predict the direction of large negative excess returns than the direction of large positive excess returns, using the level of past excess returns.

The results of $M_{ZY}(1,2)$ examines whether past volatility can be used to predict the direction of future excess returns. For all indices, the direction of the excess returns with zero threshold is not predictable using past volatility. For the excess returns with large thresholds ($c = 1, 1.5$), however, past volatility can be used to predict the direction in most cases. These results are consistent with stylized fact that there exists persistent volatility clustering for stock returns, while there exists little or weak serial dependence with very small unconditional mean.[12] Except for NASDAQ index, $M_{ZY}(1,2)$ is monotonically increasing in threshold level c. The larger the threshold, the more predictable the direction of excess returns using past volatility. There exists strong evidence that one-sided directions are easier to predict than the sign direction using past volatility. Furthermore, the direction of positive excess returns are easier to predict than that of negative excess returns, using past volatility.

The statistic $M_{ZY}(1,3)$ examines whether skewness of past excess returns is useful in predicting the direction of future excess returns. Table 6 reports the values of statistic $M_{ZY}(1,4)$, which examines whether kurtosis of past excess returns can be used to predict the direction of future excess returns. The results for $M_{ZY}(1,3)$ and $M_{ZY}(1,4)$ are similar to those for $M_{ZY}(1,2)$. For all four indices, there exists strong evidence that the direction of positive excess returns with large thresholds ($c = 1$ or 1.5) is predictable using skewness and kurtosis of past excess returns. It is easier to predict the direction of large positive returns than of large

[12] Recall that volatility clustering can generate directional predictability when $\delta = c - \mu$ is nonzero. Table 1 shows that for most stocks, the long-run average return μ is very small.

negative returns, using past skewness and kurtosis.

Finally, the statistic $M_{ZZ}(0,0)$ checks whether the direction of past excess returns can be used to predict the direction of future excess returns. This can tell us to what extent the direction of past returns contains useful information about the direction of future returns. For all four indices, the direction of future excess returns, with any threshold c, is predictable using the direction of past excess returns. The statistic $M_{ZZ}(0,0)$ is very large in most cases. This striking evidence differs from Christoffersen and Diebold's (2002) conjecture that directional dependence may not be likely to be found via analysis of directional autocorrelation for high-frequency (e.g., daily) financial data. This suggests that the directional dynamics of stock returns may be more complicated than the model considered in Christoffersen and Diebold's (2002). In general, it is easier to predict the direction of large one-sided excess returns than that of small one-sided ones, using the direction of past excess returns. And it is easier to predict the direction of large negative excess returns using the direction of past negative excess returns than to predict the direction of large positive returns using the directions of past large positive excess returns.

We have found that the directions of excess returns with any threshold is predictable. The level, volatility, skewness, and kurtosis and direction of past excess returns can be used to predict the direction of excess returns. It is well-known that there exists persistent volatility clustering in stock returns, and there may also exist weak serial dependence in the level of stock price changes, which violates the efficient market hypothesis. These well-known stylized facts may contribute to the directional predictability of stock returns. To check whether the directional predictability can be solely explained by persistent volatility clustering and mild serial dependence in mean, we fit the following MA(1)-Threshold GARCH(1,1) model via maximum likelihood estimation (MLE) to each excess stock return series:

$$\begin{cases} Y_t = \alpha_0 + \alpha_1 u_{t-1} + u_t, \\ u_t = h_t^{1/2} \varepsilon_t, \\ h_t = \beta_1 + \beta_2 h_{t-1} + u_{t-1}^2 \mathbf{1}(u_{t-1} < 0) + \beta_3^+ u_{t-1}^2 \mathbf{1}(u_{t-1} \geq 0) \\ \{\varepsilon_t\} \sim \text{i.i.d.N(0,1)}. \end{cases}$$

Here, the MA(1) component is the commonly used model to capture weak serial dependence in mean for daily stock returns. It is well-known that the GARCH model can capture persistent volatility clustering. The different coefficients β_3^- and β_3^+ allow to capture asymmetry in volatility, such as the leverage effect. This is the well-known threshold GARCH model, introduced in Glosten *et al.* (1993).

We use our tests to check the directional predictability for the fitted standardized residuals $\{\hat{\varepsilon}_t\}$, and find that the directions of sign, positive and negative $\hat{\varepsilon}$ are significantly predictable

for most thresholds c (including zero), although the test statistic values are much smaller than those based on the raw return series. This indicates the MA(1)-Threshold GARCH model cannot completely explain the directional predictability of excess stock returns.

6 Out-of-Sample Forecasts and Trading Profit

We now examine whether the documented directional predictability can be exploited to yield significant out-of-sample economic outcomes via a class of autologit forecast models. Out-of-sample evaluation is important to alleviate the problem of overfitting the data and obtain spurious results.

We use two out-of-sample evaluation measures for the directional forecast models—directional forecast accuracy and risk-adjusted profitability of model-based trading rules. For the former, we consider two statistical measures—the Quadratic Probability Score (Brier 1950, QPS) and the ratio of correct forecast directions. For the latter, we consider model-based trading rules against the most commonly used benchmark— the buy-and-hold strategy; we compare their risk-adjusted returns, including Sharpe's (1966) ratios.

6.1 Forecast models and Combined Forecasts

For comparison, we use four stock indices—DJIA, NYSE, SP500 and NASDAQ, with the same sample period from 01/02/1973 to 12/31/2001. To examine robustness of our results, we consider three sample periods—the whole sample, and two sub-sample periods: the pre-Black Monday period (from 01/02/1973 to 10/16/1987) and the post-Black Monday period (from 10/19/1987 to 12/31/2001). Each sample is divided into two subsets: the in-sample data, used to estimate model parameters; and the out-of-sample data, used to evaluate forecast performances. Table 3 lists each sample horizon and sample sizes for four stock indices.

Table 3. The horizon and total observations of three sample periods

	In-sample	Out-of-sample	Total observation (In/Out)
Whole Sample	from 01/02/1973 to 12/31/1993	from 01/03/1994 to 12/31/2001	7324 (5309/2015)
Pre-Black Monday	from 01/02/1973 to 12/30/1983	from 01/03/1984 to 10/16/1987	3738 (2779/959)
Post-Black Monday	from 10/19/1987 to 12/31/1997	from 01/02/1998 to 12/31/2001	3584 (2580/1004)

We use a class of auto-logistic models, first introduced in Cox (1958) and considered in Rydberg and Shephard (2003). The autologit model extends the logit model to allow lagged dependent variables as explanatory variables. For the zero threshold ($c = 0$), we model the binary dependent variable $Z_t(c)$ by incorporating the sources of directional predictability documented earlier. We assume that $Z_t(c)$ depends on m most recent history of some explanatory variables:

$$P\left[Z_t(c) = 1 | I_{t-1}\right] = \frac{1}{1 + \exp(\theta' X_t)}, \tag{6.1}$$

where θ is a parameter vector and $X_t \equiv (1, X_{t-1}, X_{t-2}, ..., X_{t-m})' \in I_{t-1}$. This model directly yields a probability forecast for the event $Z_t(c)$. A probability forecast for the event '$Z_t(0) = 1$' issues a likelihood that the stock price will rise, while a probability forecast for the event '$Z_t(0) = 0$' issues a likelihood that the stock price will fall.

For a nonzero threshold c, we need to define a new direction indicator $Z_t(c)$:

$$Z_t(c) = \begin{cases} 2, & \text{if } Y_t > c, \\ 1, & \text{if } -c \leq Y_t \leq c, \\ 0, & \text{if } Y_t < -c, \end{cases}$$

and use an auto-multinomial logit model:

$$P\left[Z_t(c) = s | I_{t-1}\right] = \frac{\exp(\theta_s X_t)}{\sum_{s=0}^{2} \exp(\theta_s X_t)}, \qquad s = 0, 1, 2. \tag{6.2}$$

Based on the results of directional predictability documented in Section 5, for both (6.1) and (6.2), we consider the following five models:

$$\theta_s^{(k)} X_t = \begin{cases} \theta_{s0}^{(1)} + \sum_{j=1}^{m} \theta_{sj}^{(1)} Z_{t-j} & \text{if } k = 1 \text{ (using past directions)} \\ \theta_{i0}^{(2)} + \sum_{j=1}^{m} \theta_{sj}^{(2)} Y_{t-j} & \text{if } k = 2 \text{ (using past levels)}, \\ \theta_{s0}^{(3)} + \sum_{j=1}^{m} \theta_{sj}^{(3)} Y_{t-j}^2 & \text{if } k = 3 \text{ (using past volatilities)}, \\ \theta_{s0}^{(4)} + \sum_{j=1}^{m} \theta_{sj}^{(4)} Y_{t-j}^3 & \text{if } k = 4 \text{ (using past skewness)}, \\ \theta_{s0}^{(k)} + \sum_{j=1}^{m} \theta_{sj}^{(5)} Y_{t-j}^4 & \text{if } k = 5 \text{ (using past kurtosis)}, \end{cases}$$

where $s = \{0, 1\}$ for $c = 0$ and $s = \{0, 1, 2\}$ for $c = 0.5$. All models are estimated via MLE using the in-sample observations $\{Y_t\}_{t=1}^{T_1}$ of size T_1. For each model, we first set a maximal lag order 30 for m and then use the BIC criterion to select a suitable m.

For each c, we also consider a probability forecast procedure by combining all five forecast models:

$$\hat{P}_t^{CB}(c) = \sum_{k=1}^{5} w_{kt} \hat{P}_{kt}(c),$$

where $\hat{P}_{kt}(c)$ is the 1-step-ahead probability forecast for event $Z_t(c)$ using model k, and the weights $\{w_{kt}\}$ are selected using the rule: (i) equal weighting $w_{kt} = \frac{1}{5}$ for all k and t, or (ii) time-varying weighting $w_{k,t} = \mathbf{1}[\hat{C}_{kt}(c) = 1]/\sum_{k=1}^{5} \mathbf{1}[\hat{C}_{kt}(c) = 1]$.[13] Here, $\hat{C}_{kt}(c)$ is the correct directional forecast indicator, which equals 1 when model k at $t-1$ correctly forecasts the direction of changes at t, and equals 0 otherwise. This gives a penalty when a model performs poorly. On the other hand, equal weighting is simple and most commonly used in practice. Like a portfolio, a combined forecast procedure is expected to yield more robust forecast results than a single forecast model (see Bates and Granger 1969 and Granger 2001 for more discussion.)

In order to determine $\hat{C}_{kt}(c)$, we need to use some decision rule to translate $\hat{P}_{kt}(c)$ for event $Z_t(c)$ into an event forecast. We use the following simple rule: if the forecast probability $\hat{P}_{kt}(c)$ is higher than a prespecified probability threshold, then we predict that event $Z_t(c)$ will occur. Because we consider different thresholds c, we use the in-sample proportion for the event $Z_t(c)$ as the probability threshold.

We consider two event forecasts:

$$\hat{D}^{+}_{kt}(c) = \begin{cases} 1 & \text{if } \hat{P}_k(Y_t > c|I_{t-1}) > \overline{f}^{+}(c), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\hat{D}^{-}_{kt}(c) = \begin{cases} 1 & \text{if } \hat{P}_k(Y_t < -c|I_{t-1}) > \overline{f}^{-}(c), \\ 0 & \text{otherwise,} \end{cases}$$

where $\bar{f}^{+}(c) = T_1^{-1}\sum_{t=1}^{T_1} \mathbf{1}(Y_t > c)$ and $\bar{f}^{+}(c) = T_1^{-1}\sum_{t=1}^{T_1} \mathbf{1}(Y_t < -c)$ are the in-sample proportions for the events $Z_t(c) = \mathbf{1}(Y_t > c)$ and $Z_t(c) = \mathbf{1}(Y_t < -c)$ respectively.

Given the event forecast indicators $\hat{D}^{+}_{kt}(c)$ and $\hat{D}^{-}_{kt}(c)$, we can define the correct forecast indicator

$$\hat{C}_{kt}(c) = \begin{cases} 1 & \text{if } \hat{D}^{+}_{kt}\mathbf{1}(Y_t > c) = 1 \text{ or } \hat{D}^{-}_{kt}\mathbf{1}(Y_t < -c) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The benchmark—the buy-and-hold trading strategy, is a commonly used long-term investing strategy without revising one's asset position until the end of the investment horizon. This trading strategy is equivalent to a constant probability forecast over time:

$$\hat{P}_{BH}(Y_t > c|I_{t-1}) = 1 \text{ and } \hat{P}_{BH}(Y_t < -c|I_{t-1}) = 0 \text{ for all } c, t.$$

As a result, $\hat{D}^{+}_{BHt}(c) = 1$ and $\hat{D}^{-}_{BHt}(c) = 0$ for all t and all c.

[13] When the denominator in the time-varying weighting rule is zero, (i.e., $\sum_{k=1}^{5} \mathbf{1}[\hat{C}_{kt}(c) = 1] = 0$), we use equal weighting to each model instead.

6.2 Model Evaluation Measures

We consider both statistical and economic evaluation measures for our forecast models. A popular quantitative measure for out-of-sample probability forecast accuracy is the Quadratic Probability Score (QPS), which is analogous to the Mean Squared Error:

$$\text{QPS}_k = \frac{1}{T_2} \sum_{t=T_1+1}^{T} 2[\hat{P}_{kt}(c) - Z_t(c)]^2,$$

where $\hat{P}_{kt}(c)$ is the *ex ante* probability forecast for event $Z_t(c)$ using model k, $Z_t(c)$ is the *ex post* observed direction indicator, and $T_2 \equiv T - T_1$ is the size of the out-of-sample data. QPS ranges from 0 to 2, and becomes closer to zero when a model gives a more accurate directional forecast (see Diebold and Rudebusch 1989 for more discussion). Corresponding to events $Z_t(c) = \mathbf{1}(Y_t > c)$ and $Z_t(c) = \mathbf{1}(Y_t < -c)$, we compute QPS for probability forecasts for both positive (+) and negative (-) changes. We also report directional forecast correctness ratio, which is the ratio of correct directional forecasts for event $Z_t(c)$ to the total number of occurrence of event $Z_t(c)$. Specifically, we consider positive and negative directional correctness ratios, and the overall directional correctness ratio which includes forecasts of both directions.

Our ultimate goal is to examine profitability of our forecast models. For this purpose, we define two trading rules based on model k:

$$\hat{S}_{kt}^{(1)}(c) = \begin{cases} 1 & (= \text{“buy”}) \text{ if } \hat{D}_{kt}^{+}(c) = 1, \\ -1 & (= \text{“Sell”}) \text{ if } \hat{D}_{kt}^{-}(c) = 1, \\ 0 & (= \text{“no action”}) \text{ otherwise,} \end{cases}$$

and

$$\hat{S}_{kt}^{(2)}(c) = \begin{cases} 1 & (= \text{“buy”}) \text{ if } \hat{D}_{kt}^{+} = 1, \\ 0 & (= \text{“no action”}) \text{ otherwise,} \end{cases}$$

$t = T_1 + 1, ..., T - 1$, where $\hat{D}_{kt}^{+}(c)$ and $\hat{D}_{kt}^{-}(c)$ are the event forecasts. The first trading rule allows short sales but the second one does not.[14] The out-of-sample trading return generated from model k using trading rule r is

$$\hat{R}_{kt}^{(r)} = \frac{1}{T_2} \sum_{t=T_1+1}^{T} \hat{S}_{kt}^{(r)}(c) Y_t, \qquad r = 1, 2.$$

While most studies on technical trading rules (e.g., Allen and Karjalainen 1999, Brock, LaKonishok and LeBaron 1992, Lo, Mamaysky and Wang 2000, Sullivan, Timmermann and White 1999) generally evaluate raw excess returns, Brown, Goetzmann and Kumar (1998) and Neely (2003) emphasize that it is important to consider risk-adjusted returns in comparing the

[14] Note that, for the 'buy and hold' strategy, we have $S_{BHt}^{(1)}(c) = \hat{S}_{BHt}^{(2)}(c) = 1$ for all t and c, respectively.

usefulness of trading rules, because different rules involve different levels of risk. To take this into account, we use two risk-adjusted return measures: one is the mean/standard deviation ratio, and the other is the most commonly used Sharpe's (1966) ratio, which is the unconditional expected return per unit of risk and is usually expressed in annual terms (See Sharpe 1994 for a survey):

$$\mathrm{SP}_k^{(r)} = \frac{\overline{R}_k^{(r)} - \overline{R}_f}{\hat{\sigma}_{\hat{R}_k^{(r)}}}, \qquad r = 1, 2,$$

where $\overline{R}_k^{(r)}$ is the average returns of model k using trading rule r, $\overline{R}_f$ is the average of risk-free rates, and $\hat{\sigma}_{\hat{R}_k^{(r)}}$ is the sample standard deviation of the returns $\hat{R}_k^{(r)}$. This measure is higher with a higher return and/or a lower volatility. Like most studies using the Sharpe ratio, we ignore transaction costs, which are a complicated issue.

To assess statistical significance of the difference of risk-adjusted returns between a model and the buy-and-hold strategy, we use the popular Diebold and Mariano's (1995) test. Suppose $\bar{d}_{T_2}$ denotes the difference in out-of-sample average risk-adjusted returns between a forecast model/trading strategy and the benchmark. Then under the null hypothesis that the model/strategy performs the same as the buy-and-hold strategy, Diebold and Mariano's (1995) test statistic $DM \equiv \sqrt{T_2}\,\overline{d}_{T_2}/\sqrt{2\pi \hat{f}_d(0)} \to N(0,1)$ in distribution, where $2\pi \hat{f}_d(0)$ is a Bartlett kernel-based estimator of the asymptotic variance of $\sqrt{T_2}\bar{d}_{T_2}$.

6.3 Empirical Findings

Tables 4–7 report the out-of-sample results of various forecast models for DJIA, SP500, NYSE and NASDAQ. Each table includes three sample periods—two sub-sample periods and the whole sample. The top and bottom panels in each period correspond to the results of the autologit model (for $c = 0$) and the auto-multinomial logit model (for $c = 0.5$) respectively.

The first two sections in Tables 4–7 report directional forecast accuracy measures—QPS(+), QPS(-) and forecast correctness ratios for positive direction (+), negative direction (−) and both directions (overall). In terms of QPS, all forecast models perform more or less similarly. The combined forecast models, with either time-varying weighting or equal weighting, consistently give the smallest or close to the smallest QPS for both thresholds ($c = 0, 0.5$), all the sample periods, and all stock indices. This indicates the merit of combined forecasts. Interestingly, for all forecast models, the magnitudes of QPS are closer to zero at $c = 0.5$ than at $c = 0$, except for NASDAQ in some periods. This is consistent with the earlier findings using the generalized cross-spectral tests that there is stronger evidence for directional predictability with nonzero thresholds. The QPS results also suggest that the directions of stock indices returns are more predictable during the pre-Black Monday period than the post-Black Monday

period.

In terms of directional forecast correctness ratios, the forecast model using the directions of past returns have consistently high overall correctness rations in all scenarios. The forecast model using the levels of past returns also perform well. Models using past volatilities, skewness and kurtosis do not have robust overall correctness ratios, which can be the highest or lowest among all individual forecast models. The combined forecast models generally do not have the highest correctness ratio, but they are robust and have relatively high correctness ratios. Like QPS, there is evidence that the direction of stock returns with threshold $c = 0.5$ is easier to predict than the direction of stock returns with threshold $c = 0$.

We now turn to the risk-adjusted returns—the mean/standard deviation ratio and the Sharpe ratio, which is our ultimate goal of comparing our forecast models and the buy-and-hold strategy. Consistent with the patterns for overall directional correctness ratios, the forecast models using the directions and levels of past returns have relatively robust and high risk-adjusted returns in many cases, while the forecast models using past volatilities, skewness and kurtosis do not have robust risk-adjusted returns. Interestingly, the combined forecast models, with either time-varying weighting or equal weighting, yield robust and high (sometime highest) risk-adjusted returns in all scenarios. The combined forecast model with time-varying weighting performs a bit better than the combined forecast model with equal weighting. In all scenarios the combined forecast models outperform the but-and-hold strategy, and the differences in the magnitudes of risk-adjusted returns between the combined forecast models and the buy-and-hold strategy are significantly at the 5% level in many cases.

Although the forecast model with highest directional correctness ratio does not necessarily give the highest risk-adjusted returns, the forecast models with low direction correctness ratios always give low risk-adjusted returns, and the forecast models with relatively high risk-adjusted returns usually have high directional correctness ratios. There is some evidence that directional forecast correctness and risk-adjusted returns are positively related to certain degree, indicating a positive relationship between directional forecast accuracy and performance of trading rules in many cases, as Leitch and Tanner (1991) argue. Moreover, all forecast models usually have a higher risk-adjusted returns during the pre-Black Monday period than the past-Black Monday period.

Figures 5-8 depict the out-of-sample cumulative daily returns of the combined forecast models with time-varying weighting for each threshold c, each trading rule, and each sample period, relative to the buy-and-hold strategy. These equity curves can better describe competitive positions of our model-based trading rules over time against the buy-and-hold strategy. In all scenarios, the combined forecast model with time-varying weighting achieves not only a higher cumulative return at the terminal time but also a noticeably smoother increase of

returns than the buy-and-hold strategy. The differences in cumulative returns between the combined forecast model and the buy-and-hold strategy are economically significant in many cases, particularly for the whole sample period and the post-Black Monday period. On the other hand, a roughly linear rise of the equity curve indicates that the combined forecast model can continue to earn a positive return even in a bear market, which is in sharp contrast to the buy-and-hold strategy, as can be easily seen in Figures 5-8 (E) and (F). Relatively to the buy-and-hold strategy, the combined forecast model performs better in the post-Black Monday period than the pre-Black Monday period. There is evidence of out-of-sample predictive ability of the trading rules based on directional forecast models. We also note that in most cases, the trading rule with short sales gives a higher cumulative return at the terminal time than the trading rule with no short sales. All of these graphical features are more or less consistent with the Sharp ratio results.

To sum up, we have found that combined forecast models yield robust and significantly higher risk-adjusted returns than the buy-and-hold strategy in all cases. There is significant evidence that the combined directional forecast models have some out-of-sample predictive power for the directions of stock returns.

7 Conclusion

We have proposed a rigorous model-free statistical procedure to check whether the direction of the changes of an economic time series variable is predictable using the past history of its changes. A class of separate inference procedures are also given to gauge possible sources for directional predictability. In particular, they can reveal information about whether the direction of future asset returns is predictable using the direction, level, volatility, skewness, and kurtosis of past asset returns. The proposed procedures provide reasonable references in finite samples. They have good power because they employ many lags simultaneously and discount higher order lags via the kernel function, which is consistent with the conventional wisdom that financial markets are more influenced by the recent events than by the remote events.

We have applied the proposed procedures to four daily U.S. stock price indices. We find overwhelming evidence that the direction of excess stock returns is predictable using the past history of excess stock returns. The evidence is stronger for the predictability of the directional predictability of large excess stock returns—both positive and negative. In particular, the direction and level of past excess stock returns can be used to predict the direction of future excess stock returns with both zero and nonzero thresholds, and the volatility, skewness and kurtosis of past excess stock returns can be used to predict the direction of future excess

stock returns with nonzero thresholds. The well-known weak serial dependence in mean and persistent volatility clustering in stock returns cannot explain all documented directional predictability. We finally examine the out-of-sample profitability of a class of autologit models for directional forecasts. Trading rules based on these forecast models, particularly their combinations, can earn significantly higher risk-adjusted returns than the buy-and-hold strategy.

References

Allen, F. and R. Karjalainnen (1999), "Using Generic Algorithms to Find Technical Trading Rules," *Journal of Financial Economics* 51, 245-271.

Barberis, N., A. Shleifer and R. Vishny (1998), "A Model of Investor Sentiment," *Journal of Financial Economics* 49, 307-343.

Bierens, H. (1982), "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134.

Betas, J.M. and C.W.J. Granger (1969), "The Combination of Forecasts", *Operations Research Aquarterly* 20, 451-468.

Breen, W., L.R. Glosten and R. Jagannathan (1989), "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance* 44, 1177-1189.

Brier, G.W. (1950), "Verification of Forecasts Expressed in Terms of Probability," *Monthly-Weather Review* 75, 1-3.

Brillinger, D. R. and M. Rosenblatt (1967a), "Computation and Interpretation of the kth Order Spectra," in B. Harris (ed.), *Spectral Analysis of Time Series*, Wiley: New York, pp.189-232.

Brock, W., J. Lakonishok and B. LeBaron (1992), "Simple Technical Trdaing Rules and the Stochastic Properties of Stock Returns," *Journal of Finance* 47, 1731-1764.

Brown, S., W. Goetzmann and A. Kumar (1998), "The Dow Theory: William Peter Hamilton's Track Record Reconsidered," *Journal of Finance* 53, 1311-1333.

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

Christoffersen, P. and F. Diebold (2002), "Financial Asset Returns, Market Timing, and Volatility Dynamics," Working paper, Faculty of Management, McGill University and Department of Economics, University of Pennsylvania.

Cox, D. R. (1958), "The regression analysis of binary sequences," *Journal of Royal Statistical Society* B 20, 215-242.

Cumby, R. and D. Modest (1987), "Testing for Market Timing Ability: A Framework for Forecast Evaluation," *Journal of Financial Economics* 19, 169-189.

Diebold, F. X., and Glenn D. Rudebusch (1989), "Scoring the leading indicators" *JournaI of business* 62, 369-91.

Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy" *Journal of Business and Economic Statistics* 13, 253-263.

Diebold, F.X. and J. Lopez (1996), "Forecast Evaluation and Combination," in G.S. Maddala and C.R. rao (eds.), *Handbook of Statistics*, North-Holland: Amsterdam, pp. 241-268.

Drost, F. C. and T. E. Nijman (1993), "Temporal Aggregation of GARCH Processes," *Econometrica* 61, 909-927.

Durlauf, S. (1991), "Spectral Based Tests for the Martingale Hypothesis", *Journal of Econometrics,* 50, 1-19.

Engle, C. (1994), "Can the Markov Switching Model Forecast Exchange Rates," *Journal of International Economics 36, 151-165.*

Engle, C. and J. Hamilton (1989), "Long Swings in the Dollar: Are They in the Data and Do Markets Know it?," *American Economic Review* 80, 689-713.

Fair, R.C. (1993), "Estimating Even Probabilities from Macroeconometric Models Using Stochastic Simulations," in Chapter 3, *Business Cycles, Indicators, and Forecasting,* (eds.) J. Stock and M. Watson, University of Chicago Press: Chicago.

Fama, E. F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383-417.

Fama, E. (1991), "Efficient Capital Markets: II," *Journal of Finance* 46, 1575-1618.

Glosten, L., R. Jagannathan and D. Runkle (1993), "On the Relation Between the Expected Value and Volatility of the Nominal Excess return pm Shocks," *Journal of Finance* 41, 465-474.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica 37*, 424-438.

Granger, C.W.J. (1980), "Testing for Causality: A Personal View," *Journal of Economic Dynamics and Control* 2, 329-352.

Granger, C.W.J. and M.H. Pesaran (1999), "A Decision Theoretic Approach to Forecast Evaluation," Working paper, Department of Economics, University of California, San Diego.

Hansen, B. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review* 35, 705-730.

Harvey, C. and A. Siddique (2000), Conditional Skewness in Asset Pricing Tests," *Journal of Finance* 55, 1263-1296.

Henriksson, R.D. and R. Merton (1981), "On Market Timing and Investment Performance. II: Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business* 54, 513-533.

Hong, H. and J. Stein (1999), "A Unified Theory of Underreaction, Momentum Trading and Overeaction in Asset Markets," *Journal of Finance* 54, 2143-2184.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach", *Journal of the American Statistical Association*, 84, 1201-1220.

Hong, Y. and T. Lee (2003), "Diagnostic Checking for Adequacy of Nonlinear Time Series Models," Forthcoming in *Econometric Theory.*

Jiang, G. and J. Knight (2002), "Estimation of Continuous Time Processes via the Empirical Characteristic Function," *Journal of Business and Economic Statistics,* 20, 198-212.

Knight, J. and J. Yu (2002), "Empirical Characteristic Function in Time Series Estimation," Forthcoming in *Econometric Theory.*

Kuan, C.M. and T. Liu (1995), " Forecasting Exchange Rates Using Feed-forward and Recerrent Neural Networks," *Journal of Applied Econometrics* 10, 3470264.

Lam, K. and K.C. Lam (2000), "Forecasting for the Generation of Trading Signals in Financial Markets," *Journal of Forecasting* 19, 39-52.

Leitch, G. and J. Tanner (1991), "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review* 81, 580-590.

Leitch, G. and J. Tanner (1991), "Professional Economic Forecasts: Are They Worth Their Costs," *Journal of Forecasting* 14, 143-157.

Lequarre, J.Y. (1993), "Foreign Currency Dealing: A Brief Introduction, in A.S. Weigend and N.A. Gershenfeld (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley.

Lo, A. and A. C. MacKinlay (1988), "Stock Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies* 1, 41-66.

Lo, A. and A. C. MacKinlay (1999), *A Non-Random Walk Down Wall Street.* Princeton University Press: Princeton, New Jersey.

Lo, A., H. Mamaysky and J. Wang (2000), "Foundations of Technical Analysis: Computaional Algorithms, Statistical Inference, and Empirical Implementation," *Journal of Finance* 55, 1705-1770.

Lucas, R.E. (1978), "Asset Prices in an Exchange Economy", *Econometrica,* 46, 1429-1446.

Mcqueen, Pinegar and Thorley (1996), "Delayed Reaction to Good News and the Cross-Autocorrelation of Portfolio Returns," *Journal of Finance* LI, 889-919.

Merton, R. (1981), "On Market Timing and Investment Performance: An Equilibrium Theory of Value for Market Forecasts," *Journal of Business* 54, 363-406.

Neely, C.J. (1993), "Risk-Adjusted, Ex Ante, Optimal Technical Trading Rules in Equity Markets," *International Review of Economics and Finance* 12, 69-87.

Newey, W. and K. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Pagan, A. and W. Schwert (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics* 84, 205-231.

Pesaran, M.H. and A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance," *Journal of Business and Economic Statistics* 10, 461-465.

Pesaran, M.H. and A. Timmermann (1995), "Predictability of Stock returns: Robustness and Economic Significance," *Journal of Finance* 50, 1201-1228.

Pesaran, M.H. and A. Timmermann (2000), "Market Timing and Return Prediction Under Model Uncertainty," Working Paper, Department of Economics, University of California, San Diego.

Pinkse, J. (1998), "Consistent Nonparametric Testing for Serial Independence," *Journal of Econometrics* 84, 205-231.

Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press: London.

Rydberg, T. H. and N. Shephard (2003), "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models," Forthcoming in *Journal of Financial Econometrics*.

Satchell, S. and A. Timmermann (1995), "An Assessment of the Economic Value of Non-Linear Foreign Exchange Rate Forecasts," *Journal of Forecasting* 15, 1995.

Sharpe, William F (1966), "Mutual Fund Performance," *Journal of Business* 39, 119-38.

Sharpe, William F (1994), "The Sharpe Ratio," *Journal of Portfolio Management* Fall, 49-58.

Singleton, K. (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," Forthcoming in *Journal of Econometrics*.

Stinchcombe M. and H. White (1998), "Consistent Specification Testing with Nuisance Parameters Present only Under the Alternative," *Econometric Theory* 14, 295-324.

Table 1. Summary statistics for sample data.

Tickets	Description	Starting date	Obs.	Mean	Std. dev.	Skew.	Kurt.	$r_y(1)$	$r_{y^2}(1)$
Stock Index									
DJIA	Dow Jones Avg.	1962/02/01	10,047	0.002	0.962	-1.971	58.123	0.079	0.095
SP500	S&P 500	1962/02/01	10,047	0.004	0.932	-1.621	44.407	0.090	0.120
NYSE	NYSE Comp	1966/10/28	8,851	0.005	0.889	-1.732	45.119	0.140	0.092
NADQ	Nasdaq Comp	1973/01/02	7,324	0.027	1.393	-0.359	13.452	0.081	0.312

Notes: (1) Starting date, the first available date which stock price is continously recorded in data files. Ending date for all indices and stocks is Dec 31, 2001. Obs., sample size(T), No change is that the closing price of sample data at t does not change from the closing price at $t-1$. This is represented as the percentage of T.
(2) Std. Dev., Standard deviation, Skew., Skewness, Kurt., Kurtosis.
(3) $\widehat{r_y}(1)$ is the first order sample autocorrelation in change of excess returns. $\widehat{r_{y^2}}(1)$ is the first order sample autocorrelation in square of excess returns.

TABLE 2 GS test statistics on excess return ($p = 21$)

Tickets	Two-sided direction				One-sided (Positive)				One-sided (Negative)			
	$c=0$	$c=0.5$	$c=1.0$	$c=1.5$	$c=0$	$c=0.5$	$c=1.0$	$c=1.5$	$c=0$	$c=0.5$	$c=1.0$	$c=1.5$
						$\mathbf{M}_{zx}(\mathbf{1},\mathbf{0})$						
DJIA	16.21	31.90	24.79	19.76	16.21	53.52	111.67	111.63	16.20	64.43	98.84	103.14
SP500	26.98	45.16	31.29	28.12	26.93	94.71	157.92	172.12	27.01	91.61	136.93	145.58
NYSE	38.89	62.19	45.39	32.64	38.82	59.66	105.71	121.46	38.96	73.56	98.93	97.62
NADQ	80.52	87.55	42.54	16.16	81.13	203.94	421.59	448.90	80.65	193.52	332.79	373.95
						$\mathbf{M}_{zx}(\mathbf{1},\mathbf{1})$						
DJIA	10.66	24.79	19.39	16.90	10.66	8.14	8.84	18.92	10.66	19.68	24.66	21.83
SP500	19.60	33.31	22.67	24.04	19.58	13.23	13.29	25.22	19.61	26.57	30.46	35.40
NYSE	34.78	54.70	44.59	34.32	34.73	18.98	17.11	21.61	34.82	46.70	43.94	43.81
NADQ	55.29	62.23	43.03	21.27	55.60	13.15	10.07	26.47	55.33	59.56	58.19	45.48
						$\mathbf{M}_{zx}(\mathbf{1},\mathbf{2})$						
DJIA	0.12	1.69	7.10	20.31	0.12	9.75	36.02	65.24	0.12	3.60	17.31	29.29
SP500	0.06	2.37	7.55	21.48	0.07	15.33	51.11	95.33	0.06	6.40	23.76	50.43
NYSE	-0.01	1.80	6.72	19.48	-0.01	10.14	40.74	82.49	-0.01	3.11	17.73	35.72
NADQ	2.82	1.01	4.56	7.09	2.80	100.79	391.50	628.33	2.81	112.89	312.67	466.47
						$\mathbf{M}_{zx}(\mathbf{1},\mathbf{3})$						
DJIA	0.03	1.19	6.60	20.06	0.03	1.08	5.50	17.02	0.03	-0.07	1.64	6.06
SP500	0.10	1.40	7.00	22.37	0.10	1.23	5.94	18.31	0.10	-0.02	1.58	7.10
NYSE	-0.09	1.01	6.12	20.60	-0.09	1.13	5.63	18.58	-0.09	-0.14	1.56	6.90
NADQ	0.43	2.38	7.97	17.91	0.44	0.58	3.11	10.64	0.43	2.51	5.83	7.89
						$\mathbf{M}_{zx}(\mathbf{1},\mathbf{4})$						
DJIA	-0.01	1.32	6.70	19.84	-0.01	1.09	5.48	16.43	-0.01	0.03	1.92	6.68
SP500	-0.03	1.36	6.87	22.51	-0.03	1.12	5.73	17.80	-0.03	0.06	1.80	7.81
NYSE	-0.05	1.24	6.47	21.47	-0.05	1.12	5.70	18.95	-0.05	0.06	1.78	7.36
NADQ	1.27	2.04	9.62	15.39	1.28	9.16	59.25	126.12	1.27	14.87	46.46	75.32
						$\mathbf{M}_{zz}(\mathbf{0},\mathbf{0})$						
DJIA	12.54	71.53	294.03	342.53	12.52	14.56	51.97	52.54	12.56	49.95	113.82	122.26
SP500	28.34	126.01	398.93	505.71	28.29	29.25	64.75	74.02	28.38	70.28	139.37	190.16
NYSE	42.43	71.17	250.18	352.49	42.31	21.19	43.00	64.56	42.54	63.08	101.60	134.59
NADQ	78.85	319.96	1455.98	2031.23	82.48	154.65	442.35	523.61	78.94	184.31	399.73	525.54

Notes : (1) GS tests are asymptotically one-sided $N(0,1)$ tests and thus upper-tailed asymptotic critical values may also be used, which are 1.65 and 2.33 at the 5% and 1% levels, respectively. $M_{zx}(1,l)$ represents test statistics on the martingale test, the serial correlation test, ARCH-in-mean test, Skewness-in-mean test and Kurtosis-in-mean test for $l = 1, ...4$, respectively. $M_{zz}(0,0)$ represents test statistics on I.I.D test.

(2) Excess return is defined by $100 \times ln(P_t/P_{t-1}) - r_t$.

(3) A preliminary bandwidth, p is crucial to run GS tests. We have computed GS test statistics for $p = 11, \ldots, 60$, but reported only for the value of $p = 21$ to save space.

(4) A threshold value c is introduced to forecast bigger changes. Higher threshold value of c implies bigger change in price.

Table 4. Evaluation Measures of Directional Probability for Dow Jones
Whole sample period (1973.1.2-2001.12.31) (c = 0.0)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2589	0.2575	1.0000	0.0000	0.5306	0.4512	0.7391	0.4512	0.7391
Autologistic with									
Sign	0.1372	0.1371	0.5657	0.4571	0.5147	0.4681	0.7560	0.6248	1.0160
Level	0.1378	0.1377	0.5235	0.5101	0.5172	0.9042	1.1927	0.9349**	1.3327**
Volatility	0.1373	0.1373	0.7598	0.2556	0.5231	0.3482	0.6360	0.4891	0.8413
Skewness	0.1370	0.1370	0.7683	0.2630	0.5311	0.4975	0.7854	0.5622	0.9034
Kurtosis	0.1378	0.1377	0.9287	0.0901	0.5351	0.4792	0.7670	0.4973	0.8050
Combined Prob. With									
Time varying weighting	0.1370	0.1370	0.5788	0.4496	0.5182	0.9973	1.3108	1.0416***	1.4477***
Equal weighting	0.1371	0.1370	0.5553	0.4719	0.5162	0.6853	0.9735	0.8002*	1.2059**

Whole sample period (1973.1.2-2001.12.31) (c = 0.5)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3859	0.1305	1.0000	0.0000	0.5570	0.4512	0.7391	0.4512	0.7391
Automultinomial with									
Sign	0.1166	0.1001	0.3095	0.7364	0.4986	0.2310	0.5187	0.6839	1.2609*
Level	0.1165	0.0993	0.5474	0.5105	0.5310	0.6317	0.9272	0.9116**	1.3580**
Volatility	0.1147	0.0999	0.4576	0.4268	0.4439	-0.0695	0.5909	0.4933	0.8787
Skewness	0.1160	0.0999	0.3810	0.3410	0.3633	-0.5188**	-0.1881**	-0.1140**	0.3389*
Kurtosis	0.1159	0.0995	0.2130	0.1674	0.1928	-0.7992***	0.4652	0.0648	0.6373
Combined Prob. With									
Time varying weighting	0.1157	0.0993	0.3710	0.6967	0.5153	0.7900	1.1409	1.1851**	1.6717***
Equal weighting	0.1157	0.0993	0.3894	0.7092	0.5310	0.6556	0.9977	1.0806**	1.5500***

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the whole sample period, Dow Jones Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5100, 0.4900)$ when $c = 0$ and $(0.2624, 0.2519)$ when c=0.5.
(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 4. (Continued) Evaluation Measures of Directional Probability for Dow Jones

Pre Black Monday (1973.1.2-1987.10.16) (c = 0.0)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2456	0.2434	1.0000	0.0000	0.5231	0.6881	1.1986	0.6881	1.1986
Autologistic with									
Sign	0.1291	0.1290	0.6152	0.4110	0.5178	0.7754	1.2859	0.9427	1.6006
Level	0.1282	0.1281	0.5351	0.5187	0.5273	0.5812	1.0911	0.8928	1.6109
Volatility	0.1288	0.1289	0.5912	0.3473	0.4748	-0.3297*	0.1790*	0.2281*	0.8792
Skewness	0.1280	0.1281	0.7234	0.3209	0.5314	0.8513	1.3623	0.9214	1.5332
Kurtosis	0.1283	0.1283	0.8717	0.1648	0.5346	0.7570	1.2678	0.7794	1.3303
Combined Prob. With									
Time varying weighting	0.1283	0.1282	0.5531	0.3846	0.4727	0.9924	1.5561	1.0742	1.7999*
Equal weighting	0.1281	0.1280	0.6052	0.4527	0.5325	0.9101	1.4208	1.0507	1.7225*

Pre Black Monday (1973.1.2-1987.10.16) (c = 0.5)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3735	0.1150	1.0000	0.0000	0.5474	0.6881	1.1986	0.6881	1.1986
Automultinomial with									
Sign	0.1021	0.0912	0.9462	0.2326	0.6232	0.6992	1.2502	0.9226*	1.4601**
Level	0.1013	0.0924	0.6346	0.4651	0.5579	0.3303	0.8517	0.7698	1.4260
Volatility	0.1015	0.0923	0.9346	0.2000	0.6021	0.6609	1.2055	0.8114	1.3406
Skewness	0.1013	0.0917	0.4769	0.4279	0.4547	-0.2173	0.3375	0.3119	1.1067
Kurtosis	0.1016	0.0926	0.9423	0.1349	0.5768	0.7643	1.2963	0.7235	1.2495
Combined Prob. With									
Time varying weighting	0.1015	0.0920	0.9269	0.2372	0.6147	0.9767	1.5412	1.0986***	1.6454***
Equal weighting	0.1015	0.0918	0.9192	0.2093	0.5979	0.8918	1.4424	1.0019**	1.5482**

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Pre Black Monday period, Dow Jones Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.4941, 0.5059)$ when $c = 0$ and $(0.2854, 0.3001)$ when $c = 0.5\widehat{\mathbf{S}}_{\mathbf{y}}$.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 4. (Continued) Evaluation Measures of Directional Probability for Dow Jones

Post Black Monday (1987.10.19-2001.12.31) (c = 0.0)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2740	0.2729	1.0000	0.0000	0.5115	0.0779	0.3155	0.0779	0.3155
Autologistic with									
Sign	0.1397	0.1397	0.5234	0.5419	0.5325	0.8000	1.0380	0.5857*	0.9032**
Level	0.1402	0.1401	0.5293	0.5419	0.5355	0.6821	0.9200	0.4906*	0.7977*
Volatility	0.1411	0.1410	0.4414	0.6074	0.5225	0.5781	0.8159	0.5196	0.8962*
Skewness	0.1404	0.1403	0.4902	0.5624	0.5255	0.8964	1.1347	0.6825**	1.0163**
Kurtosis	0.1419	0.1418	0.2949	0.6933	0.4895	0.0736	0.3111	0.1270	0.5253
Combined Prob. With									
Time varying weighting	0.1396	0.1404	0.5430	0.5358	0.5395	1.2196**	1.4841**	0.6741**	0.9821**
Equal weighting	0.1397	0.1396	0.5488	0.5297	0.5395	1.2656**	1.5042**	0.8703***	1.1794***

Post Black Monday (1987.10.19-2001.12.31) (c = 0.5)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3672	0.1735	1.0000	0.0000	0.5259	0.0779	0.3155	0.0779	0.3155
Automultinomial with									
Sign	0.1277	0.1244	0.9217	0.5916	0.7652	0.3167	0.6467	0.2178	0.4707
Level	0.1289	0.1253	0.5014	0.4920	0.4970	0.4983	0.7403	0.3552	0.7041
Volatility	0.1316	0.1247	0.5188	0.4823	0.5015	0.1525	0.5276	0.1430	0.4798
Skewness	0.1293	0.1250	0.1217	0.1897	0.1540	0.0179	0.4472	-0.0230	0.6723
Kurtosis	0.1312	0.1246	0.1768	0.2508	0.2119	-0.0587	0.3021	-0.0773	0.4995
Combined Prob. With									
Time varying weighting	0.1287	0.1244	0.7014	0.5145	0.6128	0.7581	1.0427	0.4192	0.7188
Equal weighting	0.1291	0.1243	0.6783	0.5209	0.6037	0.8055	1.0788	0.3727	0.6774

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Post Black Monday period, Dow Jones Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5384, 0.4616)$ when c=0 and (0.2729,0.2186) when c=0.5.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 5. Evaluation Measure of Directional Probability for S&P500
Whole sample period (1973.1.2-2001.12.31) (c = 0.0)

	QPS		Correctness			Trading Rule1		Trading Rule2	
Estimation Models	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2567	0.2565	1.0000	0.0000	0.5335	0.3815	0.6638	0.3815	0.6638
Autologistic with									
Sign	0.1371	0.1371	0.5447	0.5122	0.5296	0.8936	1.1764	0.9488**	1.3694***
Level	0.1380	0.1380	0.5261	0.5389	0.5320	1.1736**	1.4569**	1.1255***	1.5353***
Volatility	0.1374	0.1374	0.6201	0.4100	0.5221	0.1279	0.4101	0.3466	0.7307
Skewness	0.1381	0.1381	0.6769	0.3344	0.5171	0.1599	0.4421	0.3413	0.6972
Kurtosis	0.1377	0.1377	0.8799	0.1448	0.5370	0.3083	0.5905	0.3820	0.6947
Combined Prob. With									
Time varying weighting	0.1367	0.1371	0.5149	0.4281	0.4744	0.8648	1.1794	0.8313*	1.2515**
Equal weighting	0.1370	0.1370	0.5484	0.5176	0.5340	0.8659	1.1487	0.8763**	1.2734**

Whole sample period (1973.1.2-2001.12.31) (c = 0.5)

	QPS		Correctness			Trading Rule1		Trading Rule2	
Estimation Models	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3837	0.1360	1.0000	0.0000	0.5501	0.3815	0.6638	0.3815	0.6638
Automultinomial with									
Sign	0.1170	0.1021	0.3005	0.7369	0.4968	0.4374	0.7220	0.7399	1.2988**
Level	0.1176	0.1032	0.5468	0.5060	0.5285	0.8175	1.1162	0.9415**	1.3725***
Volatility	0.1151	0.1019	0.5008	0.4679	0.4860	-0.5533**	0.1553	0.3620	0.7329
Skewness	0.1168	0.1025	0.8670	0.8614	0.8645	-0.3362*	0.2003	0.2080*	0.5089
Kurtosis	0.1168	0.1024	0.9885	0.1687	0.6197	0.1960	0.5112	0.2860**	0.5708**
Combined Prob. With									
Time varying weighting	0.1164	0.1021	0.4581	0.6546	0.5465	0.8068	1.1418	1.2337***	1.6815***
Equal weighting	0.1163	0.1019	0.4778	0.6827	0.5700	0.9602	1.2873	1.3237***	1.7517***

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the whole sample period, S&P500 Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5139, 0.4861)$ when c=0 and (0.2649,0.2479) when c=0.5.
(2) **:** ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 5. (Continued) Evaluation Measure of Directional Probability for S&P500

Pre Black Monday (1973.1.2-1987.10.16) (c = 0.0)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2424	0.2413	1.0000	0.0000	0.5282	0.6611	1.1998	0.6611	1.1998
Autologistic with									
Sign	0.1295	0.1295	0.5366	0.4856	0.5126	0.5338	1.0720	0.8272	1.5728
Level	0.1279	0.1279	0.5485	0.4989	0.5251	0.9018	1.4409	1.0612	1.7932**
Volatility	0.1278	0.1278	0.8099	0.2395	0.5408	1.1201	1.6604	0.9896*	1.5894**
Skewness	0.1273	0.1273	0.7208	0.3304	0.5366	0.9573	1.4970	0.9562	1.5936*
Kurtosis	0.1282	0.1282	0.9030	0.1153	0.5314	0.6860	1.2248	0.7070	1.2726
Combined Prob. With									
Time varying weighting	0.1277	0.1278	0.5327	0.4523	0.4948	0.6510	1.2041	0.8429	1.5971
Equal weighting	0.1275	0.1275	0.5545	0.4767	0.5178	0.5803	1.1186	0.8587	1.6038

Pre Black Monday (1973.1.2-1987.10.16) (c = 0.5)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3724	0.1166	1.0000	0.0000	0.5458	0.6611	1.1998	0.6611	1.1998
Automultinomial with									
Sign	0.1022	0.0907	1.0000	0.0734	0.5792	0.9012*	1.4617**	0.6611***	1.1998***
Level	0.1022	0.0909	0.5534	0.5092	0.5333	0.5562	1.1224	0.6993	1.4434
Volatility	0.1020	0.0918	0.3740	0.2752	0.3292	-0.5530**	0.9432	-0.2467**	0.6130*
Skewness	0.1017	0.0909	0.8397	0.4541	0.6646	1.2520	1.8557	1.1376**	1.7509***
Kurtosis	0.1021	0.0918	0.9962	0.1422	0.6083	0.9191*	1.4922**	0.6427	1.1817
Combined Prob. With									
Time varying weighting	0.1018	0.0910	0.7366	0.3394	0.5563	0.8092	1.4243	0.9140	1.5679
Equal weighting	0.1018	0.0908	0.7366	0.3349	0.5542	0.8263	1.4415	0.9098	1.5637

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Pre Black Monday period, S&P500 Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.4991, 0.5009)$ when c=0 and (0.2875,0.2832) when c=0.5.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 5. (Continued) Evaluation Measure of Directional Probability for S&P500

Post Black Monday (1987.10.19-2001.12.31) (c = 0.0)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2768	0.2768	1.0000	0.0000	0.5055	-0.0125	0.2141	-0.0125	0.2141
Autologistic with									
Sign	0.1405	0.1405	0.5325	0.5383	0.5354	1.0818**	1.3092**	0.6890**	0.9819***
Level	0.1414	0.1414	0.5385	0.5403	0.5394	1.0760**	1.3032**	0.6896**	0.9845**
Volatility	0.1426	0.1426	0.6174	0.4093	0.5145	0.8494*	1.0766*	0.5296**	0.8171**
Skewness	0.1413	0.1413	0.5444	0.4899	0.5174	0.3490	0.5758	0.2168	0.5091
Kurtosis	0.1440	0.1440	0.4556	0.5504	0.5025	0.3903	0.6170	0.2889	0.6358
Combined Prob. With									
Time varying weighting	0.1410	0.1406	0.6252	0.4819	0.5543	1.1745**	1.4257**	0.6658**	0.9421***
Equal weighting	0.1405	0.1405	0.5917	0.4597	0.5264	0.8456*	1.0728*	0.5218**	0.8063**

Post Black Monday (1987.10.19-2001.12.31) (c = 0.5)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3566	0.1881	1.0000	0.0000	0.5193	-0.0125	0.2141	-0.0125	0.2141
Automultinomial with									
Sign	0.1316	0.1298	0.9286	0.9080	0.9187	0.3460	0.8998	0.1793*	0.4147*
Level	0.1332	0.1348	0.5247	0.4570	0.4922	0.0569	0.3217	0.1323	0.4264
Volatility	0.1298	0.1325	0.7527	0.6083	0.6833	0.5742	0.9804	0.5921***	0.8483***
Skewness	0.1358	0.1343	0.3709	0.3116	0.3424	-0.0062	0.3095	-0.0653	0.2956
Kurtosis	0.1337	0.1341	0.3077	0.6409	0.4679	0.3982	0.6367	0.3263	0.7166
Combined Prob. With									
Time varying weighting	0.1314	0.1321	0.7995	0.7359	0.7689	0.6129	0.9658	0.5427***	0.7931***
Equal weighting	0.1315	0.1320	0.7802	0.7033	0.7432	0.6004	0.9192	0.4413**	0.6940**

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Post Black Monday" period, S&P500 Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5438, 0.4562)$ when c=0 and (0.2698,0.2132) when c=0.5.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 6. Evaluation Measures of Directional Probability for NYSE
Whole sample period (1973.1.2-2001.12.31) (c = 0.0)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2592	0.2581	1.0000	0.0000	0.5299	0.3753	0.7037	0.3753	0.7037
Autologistic with									
Sign	0.1363	0.1362	0.5418	0.5376	0.5398	1.3345**	1.6643**	1.3216***	1.8309***
Level	0.1371	0.1370	0.5418	0.5249	0.5338	1.0456*	1.3747*	1.0477***	1.5327***
Volatility	0.1370	0.1370	0.6732	0.3608	0.5264	0.0572	0.3853	0.2761	0.6953
Skewness	0.1376	0.1376	0.7549	0.2751	0.5294	0.0592	0.3873	0.2576	0.6469
Kurtosis	0.1373	0.1372	0.9183	0.0921	0.5299	0.2971	0.6254	0.3626	0.7166
Combined Prob. With									
Time varying weighting	0.1366	0.1364	0.5136	0.4582	0.4876	1.0327*	1.3855*	1.0631***	1.5975***
Equal weighting	0.1364	0.1364	0.5624	0.5111	0.5383	1.1364*	1.4659*	1.1460***	1.6449***

Whole sample period (1973.1.2-2001.12.31) (c = 0.5)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3960	0.1229	1.0000	0.0000	0.5562	0.3753	0.7037	0.3753	0.7037
Automultinomial with									
Sign	0.1123	0.0970	0.2376	0.7867	0.4813	0.5949	0.9347	1.1316**	1.8926***
Level	0.1117	0.0960	0.5142	0.5089	0.5118	0.7820	1.1388	0.7885*	1.2703**
Volatility	0.1101	0.0953	0.3830	0.3822	0.3826	-0.3928*	0.7170	0.2339	0.7188
Skewness	0.1111	0.0958	0.8209	0.6378	0.7396	-0.0609	0.3749	0.3578	0.7236
Kurtosis	0.1108	0.0955	0.1862	0.1933	0.1893	-1.5082***	-0.3355**	-0.3145**	0.3462
Combined Prob. With									
Time varying weighting	0.1107	0.0953	0.2926	0.7756	0.5069	0.5159	0.8739	0.8778*	1.5435***
Equal weighting	0.1107	0.0953	0.3191	0.8022	0.5335	0.6247	0.9677	0.9929**	1.5949***

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the whole sample period, NYSE index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5168, 0.4832)$ when $c = 0$ and $(0.2645, 0.2406)$ when c=0.5.
(2) **:** ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 6. (Continued) Evaluation Measures of Directional Probability for NYSE

Pre Black Monday (1973.1.2-1987.10.16) (c = 0.0)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2381	0.2354	1.0000	0.0000	0.5383	0.6595	1.2433	0.6595	1.2433
Autologistic with									
Sign	0.1280	0.1281	0.5692	0.4841	0.5299	0.9385	1.5229	1.0956	1.8969**
Level	0.1269	0.1270	0.5497	0.4977	0.5257	0.7479	1.3317	0.9781	1.7899*
Volatility	0.1279	0.1279	0.8129	0.2045	0.5320	0.5003	1.0836	0.6506	1.3056
Skewness	0.1273	0.1272	0.6667	0.3955	0.5414	1.2613	1.8480	1.2226*	1.9686**
Kurtosis	0.1286	0.1286	0.9006	0.1273	0.5435	0.7357	1.3198	0.7406	1.3606
Combined Prob. With									
Time varying weighting	0.1273	0.1270	0.5361	0.4364	0.4900	0.8783	1.4930	0.9812	1.8037*
Equal weighting	0.1270	0.1270	0.5731	0.4818	0.5310	0.8217	1.4057	1.0074	1.8018*

Pre Black Monday (1973.1.2-1987.10.16) (c = 0.5)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3777	0.1134	1.0000	0.0000	0.5431	0.6595	1.2433	0.6595	1.2433
Automultinomial with									
Sign	0.0995	0.0903	0.2897	1.0000	0.6142	-1.6105**	-0.9254**	-0.2452**	0.8645
Level	0.1000	0.0884	0.5595	0.5425	0.5517	0.9948	1.6056	1.0583	1.8750*
Volatility	0.0997	0.0897	1.0000	0.2217	0.6444	1.0047*	1.6547**	0.6623	1.2461
Skewness	0.0996	0.0889	0.6944	0.4528	0.5841	1.1772	1.7764	0.9793	1.7017*
Kurtosis	0.0998	0.0900	0.9960	0.1085	0.5905	0.8408	1.4551*	0.6419	1.2261
Combined Prob. With									
Time varying weighting	0.0994	0.0889	0.5714	0.5660	0.5690	1.5393	2.1666	1.3741**	2.1988***
Equal weighting	0.0995	0.0889	0.5794	0.5660	0.5733	1.5452	2.1705	1.3440**	2.1625***

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Pre Black Monday period, NYSE index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.4977, 0.5023)$ when $c = 0$ and $(0.2872, 0.2735)$ when c=0.5.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 6. (Continued) Evaluation Measures of Directional Probability for NYSE

Post Black Monday (1987.10.19-2001.12.31) (c = 0)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2785	0.2785	1.0000	0.0000	0.5025	-0.0536	0.2148	-0.0536	0.2148
Autologistic with									
Sign	0.1408	0.1408	0.5655	0.4950	0.5304	1.2917**	1.5612**	0.8119***	1.1656***
Level	0.1414	0.1414	0.5238	0.5251	0.5244	0.8305*	1.0997*	0.5041**	0.8534**
Volatility	0.1416	0.1416	0.4484	0.5812	0.5145	-0.0119	0.2565	-0.0486	0.3496
Skewness	0.1423	0.1423	0.6865	0.3086	0.4985	-0.1175	0.1509	-0.1057	0.2260
Kurtosis	0.1416	0.1416	0.3294	0.7134	0.5204	0.5162	0.7849	0.3627	0.7840
Combined Prob. With									
Time varying weighting	0.1407	0.1412	0.5536	0.4489	0.5015	1.0551**	1.3536**	0.5859**	0.9405**
Equal weighting	0.1407	0.1407	0.5992	0.4689	0.5344	1.3932**	1.6630**	0.8502***	1.1929***

Post Black Monday (1987.10.19-2001.12.31) (c = 0.5)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3694	0.1752	1.0000	0.0000	0.5206	-0.0536	0.2148	-0.0536	0.2148
Automultinomial with									
Sign	0.1292	0.1228	0.3900	0.7675	0.5710	0.6377	0.9426	0.4545*	0.8819**
Level	0.1290	0.1274	0.4399	0.5127	0.4748	0.4784	0.7963	-0.0025	0.3882
Volatility	0.1269	0.1258	0.5953	0.5510	0.5740	-0.3892	0.4257	0.6272**	0.9744**
Skewness	0.1292	0.1283	0.3548	0.3248	0.3405	0.0031	0.3309	0.5774*	1.0174**
Kurtosis	0.1281	0.1277	0.3343	0.2834	0.3099	0.1927	0.6771	0.1576	0.6043
Combined Prob. With									
Time varying weighting	0.1278	0.1257	0.5337	0.7134	0.6198	0.7592	1.1028	0.8735**	1.2435***
Equal weighting	0.1277	0.1254	0.5513	0.7452	0.6443	0.9693	1.3167*	0.8463**	1.2089***

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Post Black Monday" period, NYSE index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5457, 0.4543)$ when c=0 and (0.2667,0.2054) when c=0.5.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 7. Evaluation Measure of Directional Probability for NASDAQ
Whole sample period (1973.1.2-2001.12.31) (c = 0.0)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2455	0.2453	1.0000	0.0000	0.5539	0.2429	0.4106	0.2429	0.4106
Autologistic with									
Sign	0.1363	0.1363	0.5677	0.5334	0.5524	1.0335*	1.2016*	1.0491***	1.3253***
Level	0.1519	0.1518	0.5973	0.5156	0.5609	0.9041	1.0722	0.9027**	1.1672***
Volatility	0.1409	0.1408	0.4951	0.4844	0.4903	0.0985	0.2662	0.2634	0.5222
Skewness	0.1488	0.1488	0.6296	0.4310	0.5410	0.6954	0.8633	0.6519*	0.8851**
Kurtosis	0.1421	0.1420	0.7193	0.3007	0.5325	0.7365	0.9043	0.6463*	0.8679**
Combined Prob. With									
Time varying weighting	0.1366	0.1381	0.5211	0.4510	0.4898	1.0568*	1.2416*	1.1210***	1.4252***
Equal weighting	0.1361	0.1360	0.5776	0.5234	0.5534	0.9275*	1.0956*	0.9664***	1.2438***

Whole sample period (1973.1.2-2001.12.31) (c = 0.5)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3253	0.1723	1.0000	0.0000	0.5660	0.2429	0.4106	0.2429	0.4106
Automultinomial with									
Sign	0.1519	0.1190	0.2697	0.9319	0.5571	0.1581	0.3429	0.8098**	1.1856***
Level	0.1505	0.1310	0.5541	0.4945	0.5282	1.0480*	1.2263**	0.8526**	1.1090***
Volatility	0.1605	0.1275	0.6488	0.5658	0.6128	0.0327	0.2639	0.1535	0.3350
Skewness	0.1444	0.1276	0.4228	0.3645	0.3975	0.3780	0.6111	0.2095	0.4604
Kurtosis	0.1402	0.1226	0.8591	0.1791	0.5640	0.3938	0.5626	0.4033	0.6059
Combined Prob. With									
Time varying weighting	0.1394	0.1188	0.5784	0.7322	0.6451	0.8345	1.0592*	0.4934	0.6990*
Equal weighting	0.1373	0.1183	0.5723	0.7464	0.6479	0.7097	0.9368	0.4124	0.6164

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the whole sample period, Nasdaq Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5686, 0.4314)$ when c=0 and (0.2946,0.2182) when c=0.5.
(2) **:** ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 7.(Continued) Evaluation Measure of Directional Probability for NASDAQ

Pre Black Monday (1973.1.2-1987.10.16) ($\mathbf{c = 0.0}$)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2285	0.2279	1.0000	0.0000	0.5549	0.4203	1.1514	0.4203	1.1514
Autologistic with									
Sign	0.1193	0.1194	0.6290	0.6127	0.6217	3.8313***	4.5901***	3.0418***	4.1261***
Level	0.1185	0.1184	0.6591	0.5610	0.6155	3.7791***	4.5367***	2.9591***	4.0256***
Volatility	0.1276	0.1275	0.8663	0.1221	0.5350	-0.3686	0.3602	0.0290*	0.8674
Skewness	0.1264	0.1263	0.8983	0.1596	0.5695	1.5340***	2.2714***	1.0780***	1.8900***
Kurtosis	0.1270	0.1269	0.8569	0.1455	0.5402	0.0996	0.8297	0.2784	1.0612
Combined Prob. With									
Time varying weighting	0.1219	0.1204	0.6667	0.5516	0.6155	4.0111***	4.7893***	3.0176***	4.0672***
Equal weighting	0.1219	0.1218	0.6685	0.5657	0.6228	3.9989***	4.7604***	3.0826***	4.1430***

Pre Black Monday (1973.1.2-1987.10.16) ($\mathbf{c = 0.5\widehat{S}_y}$)

Estimation Models	QPS		Correctness			Trading Rule1		Trading Rule2	
	+	-	+	-	Overall	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R1}}$	Sharpe Ratio	$\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3735	0.0963	1.0000	0.0000	0.5909	0.4203	1.1514	0.4203	1.1514
Automultinomial with									
Sign	0.0997	0.0908	0.0000	0.9111	0.3727	0.4022	1.2805	-71.4338***	N/A
Level	0.0957	0.0735	0.6962	0.6333	0.6705	4.1458***	4.9478***	3.1800***	4.2851***
Volatility	0.1006	0.0789	0.2423	0.1778	0.2159	-2.2055***	0.1543	-0.0358	1.3785
Skewness	0.1017	0.0781	0.9385	0.1722	0.6250	1.2564***	2.0063***	0.9125***	1.6911***
Kurtosis	0.1021	0.0794	0.1077	0.0778	0.0955	-2.3378***	0.9231 -	1.1795**	0.9324
Combined Prob. With									
Time varying weighting	0.0974	0.0772	0.4423	0.9167	0.6364	2.6741**	3.4314**	2.9116***	4.2908***
Equal weighting	0.0981	0.0774	0.4231	0.9222	0.6273	2.6906**	3.4518**	2.9082***	4.3292***

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Pre Black Monday period, Nasdaq Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5703, 0.4297)$ when c=0 and (0.3080,0.2353) when c=0.5.

(2) **:** ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Table 7.(Continued) Evaluation Measure of Directional Probability for NASDAQ

Post Black Monday (1987.10.19-2001.12.31) (c = 0)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.2595	0.2595	1.0000	0.0000	0.5364	0.0316	0.1552	0.0316	0.1552
Autologistic with									
Sign	0.1418	0.1418	0.5353	0.5333	0.5344	0.7141	0.8378	0.5954*	0.7930*
Level	0.1595	0.1595	0.5539	0.5226	0.5394	0.5806	0.7042	0.4593	0.6449*
Volatility	0.1433	0.1433	0.5242	0.4753	0.5015	0.2963	0.4198	0.2219	0.3892
Skewness	0.1662	0.1662	0.5465	0.4774	0.5145	0.1627	0.2863	0.1351	0.3069
Kurtosis	0.1607	0.1607	0.4758	0.5032	0.4885	0.0405	0.1641	0.0570	0.2521
Combined Prob. With									
Time varying weighting	0.1408	0.1456	0.4740	0.4968	0.4845	0.8120	0.9531	0.8471**	1.0530**
Equal weighting	0.1420	0.1420	0.4907	0.5548	0.5204	0.5334	0.6571	0.4422	0.6358

Post Black Black Monday (1987.10.19-2001.12.31) (c = 0.5)

Estimation Models	QPS +	QPS -	Correctness +	Correctness -	Correctness Overall	Trading Rule1 Sharpe Ratio	Trading Rule1 $\frac{\text{Return}}{\sigma_{R1}}$	Trading Rule2 Sharpe Ratio	Trading Rule2 $\frac{\text{Return}}{\sigma_{R2}}$
Buy and Hold	0.3092	0.2081	1.0000	0.0000	0.5462	0.0316	0.1552	0.0316	0.1552
Automultinomial with									
Sign	0.1581	0.1356	0.2673	0.7748	0.4976	0.4884	0.6121	0.5703	0.8416*
Level	0.1599	0.1501	0.5256	0.5147	0.5207	0.4685	0.5941	0.3368	0.5346
Volatility	0.1540	0.1432	0.7617	0.7453	0.7543	0.2085	0.4632	0.2151	0.3528
Skewness	0.1657	0.1562	0.4878	0.4960	0.4915	0.5018	0.6438	0.0707	0.2697
Kurtosis	0.1577	0.1537	0.6169	0.6434	0.6290	-0.0509	0.1801	0.4097	0.5717
Combined Prob. With									
Time varying weighting	0.1540	0.1391	0.4744	0.7453	0.5973	0.5819	0.7261	0.4771	0.6773
Equal weighting	0.1529	0.1365	0.4811	0.7721	0.6131	0.6841	0.8253	0.4997	0.6998*

Notes: (1) Trading Rule 1 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$, 'sell' if $\Pr(Y_{T+t} < -c) > \overline{f}^{-}(c)$, and 'no action' otherwise; Trading Rule 2 is defined by: 'Buy' if $\Pr(Y_{T+t} > c) > \overline{f}^{+}(c)$ and 'no action' otherwise (where $\overline{f}^{+}(c)$ and $\overline{f}^{-}(c)$ refer to the sample frequency in positive changes and negative changes for in-sample period, respectively). During the Post Black Monday period, Nasdaq Index has $(\overline{f}^{+}(c), \overline{f}^{-}(c)) = (0.5733, 0.4267)$ when c=0 and (0.2961,0.2163) when c=0.5.

(2) : ***,** and * indicate significance at the 1%, 5% and 10% level, respectively for the forecast comparison statistic of Diebold and Mariano (1995) with a Bartlett kernel and a truncation lag of 10. Each model is compared with 'the buy and hold' model.

Figure 1: **Rejection Rates of GS test statistics Under DGP1** ($T = 500$)

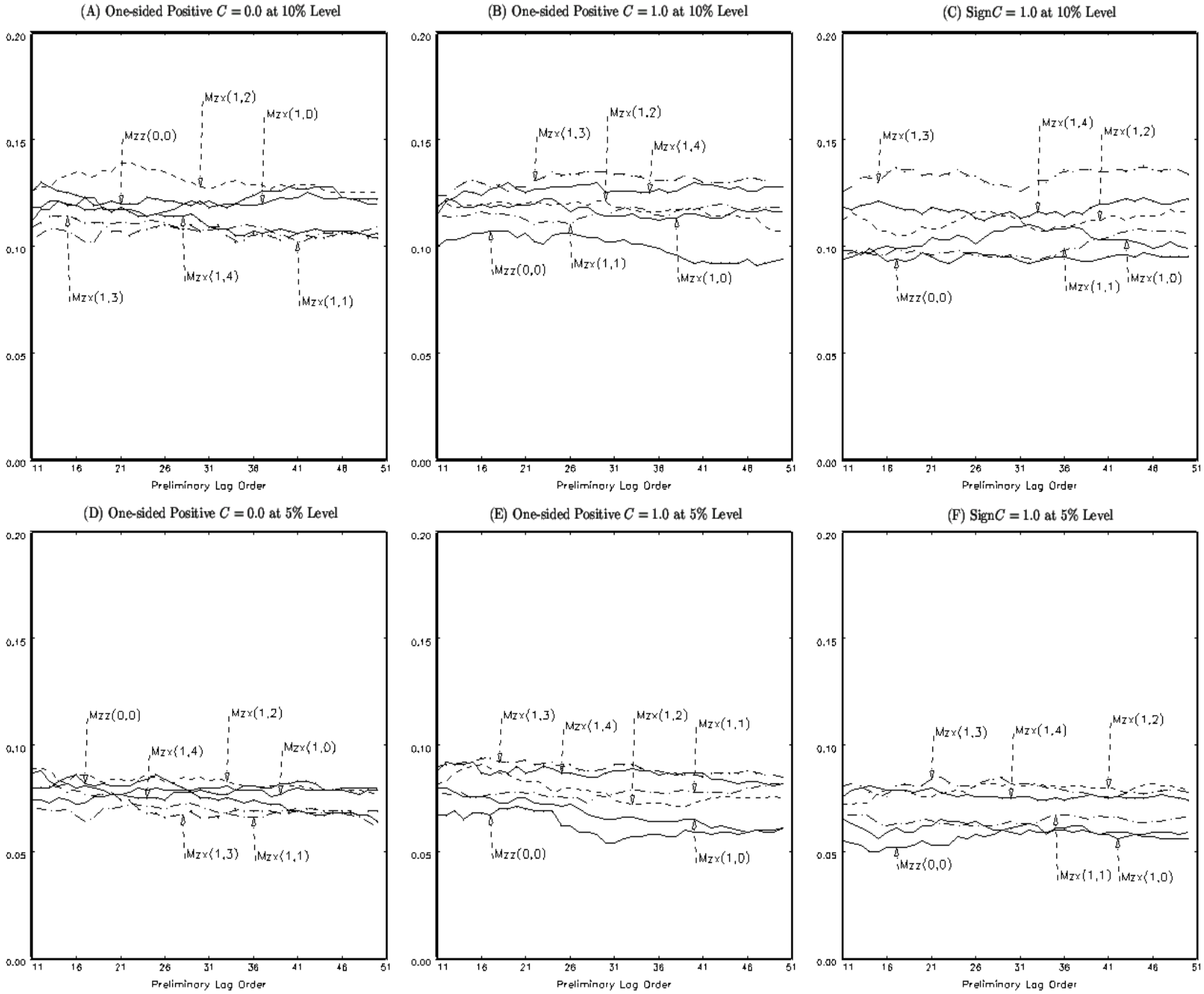

Figure 2: **Rejection Rates of GS test statistics Under DGP1** ($T = 1000$)

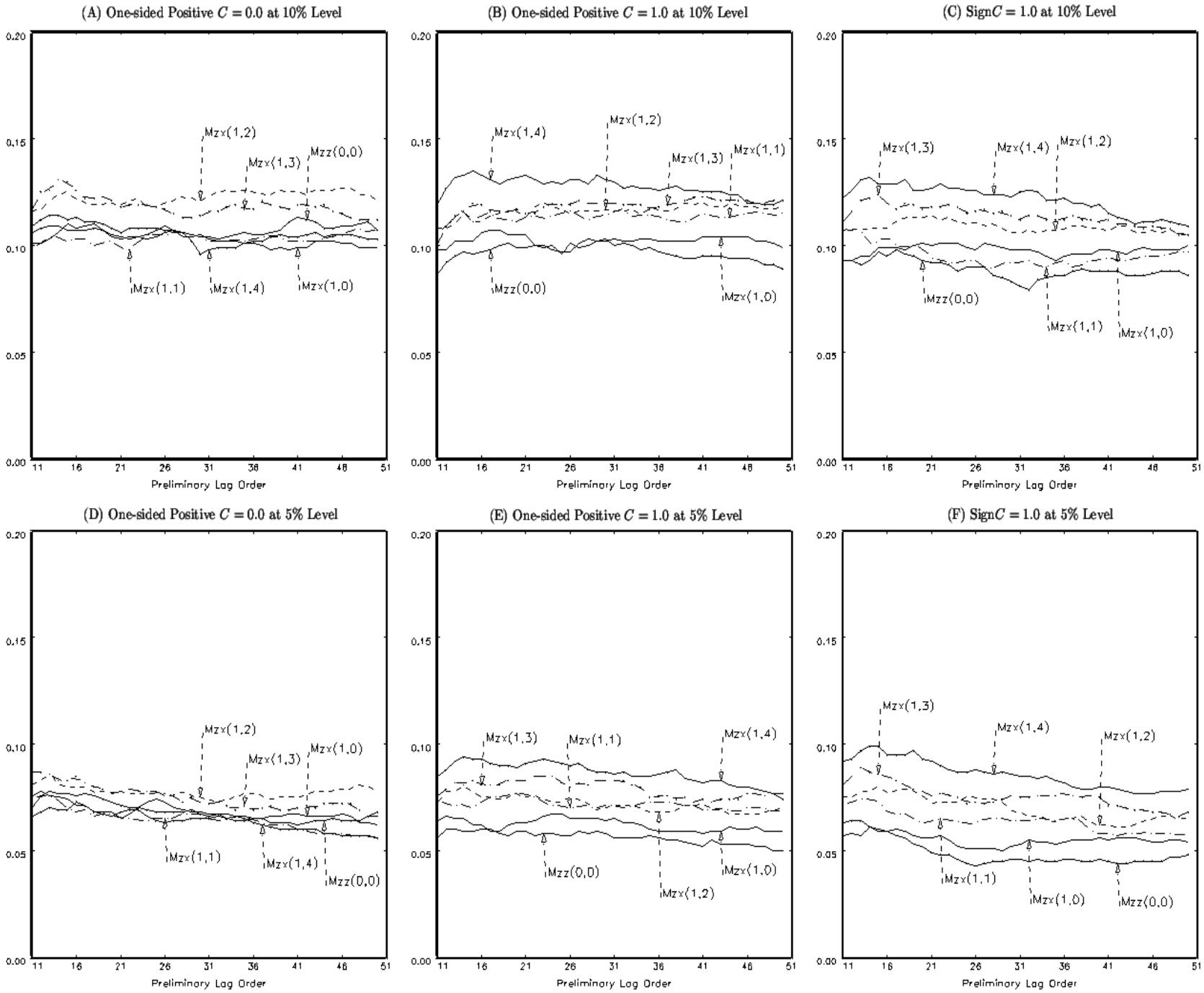

Figure 3: **Rejection Rates of GS test statistics Under DGP2** ($T = 500$)

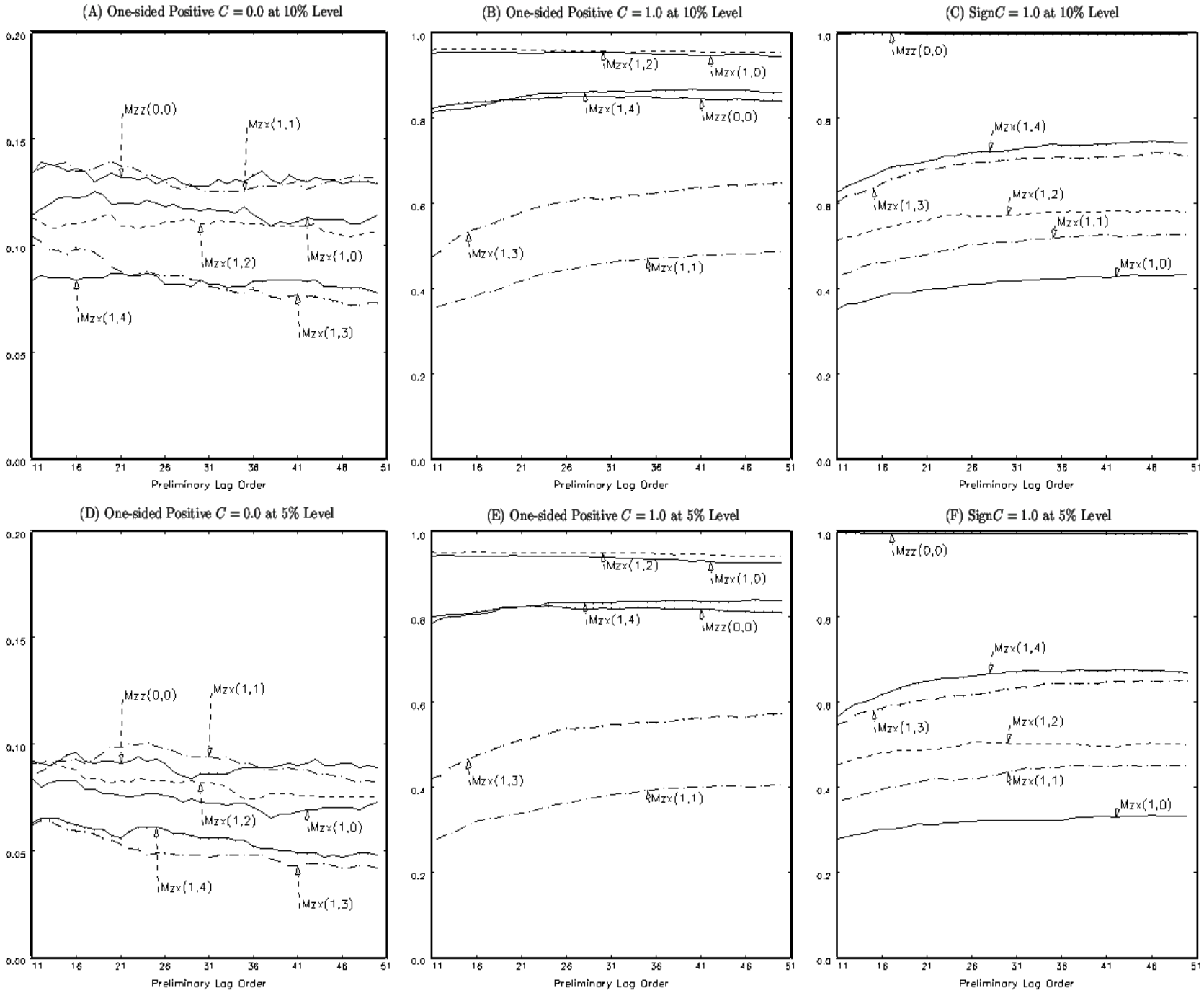

Figure 4: **Rejection Rates of GS test statistics Under DGP2** ($T = 1000$)

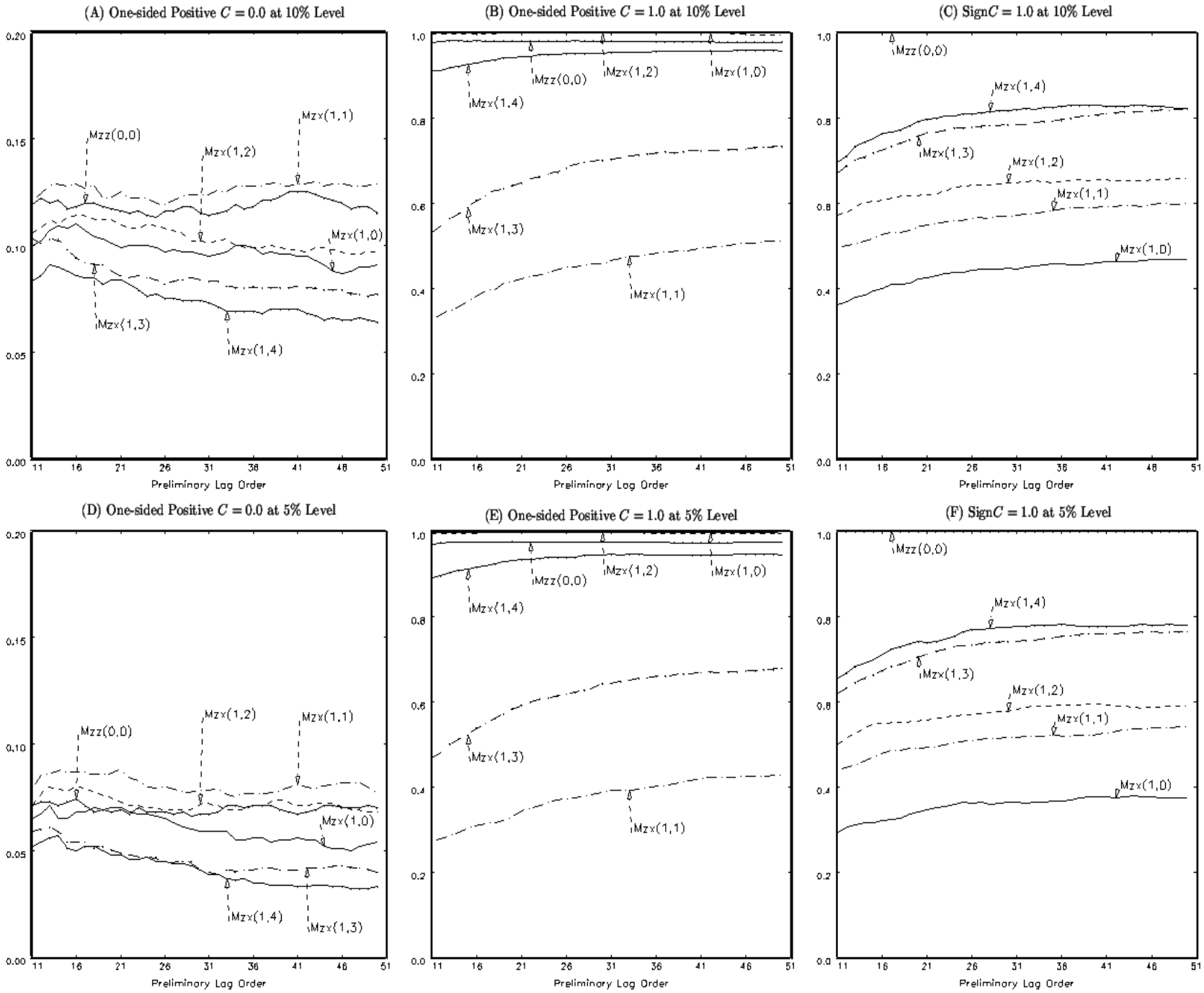

Figure 5: **Cumulative Daily Returns (%) of Dow Jones**

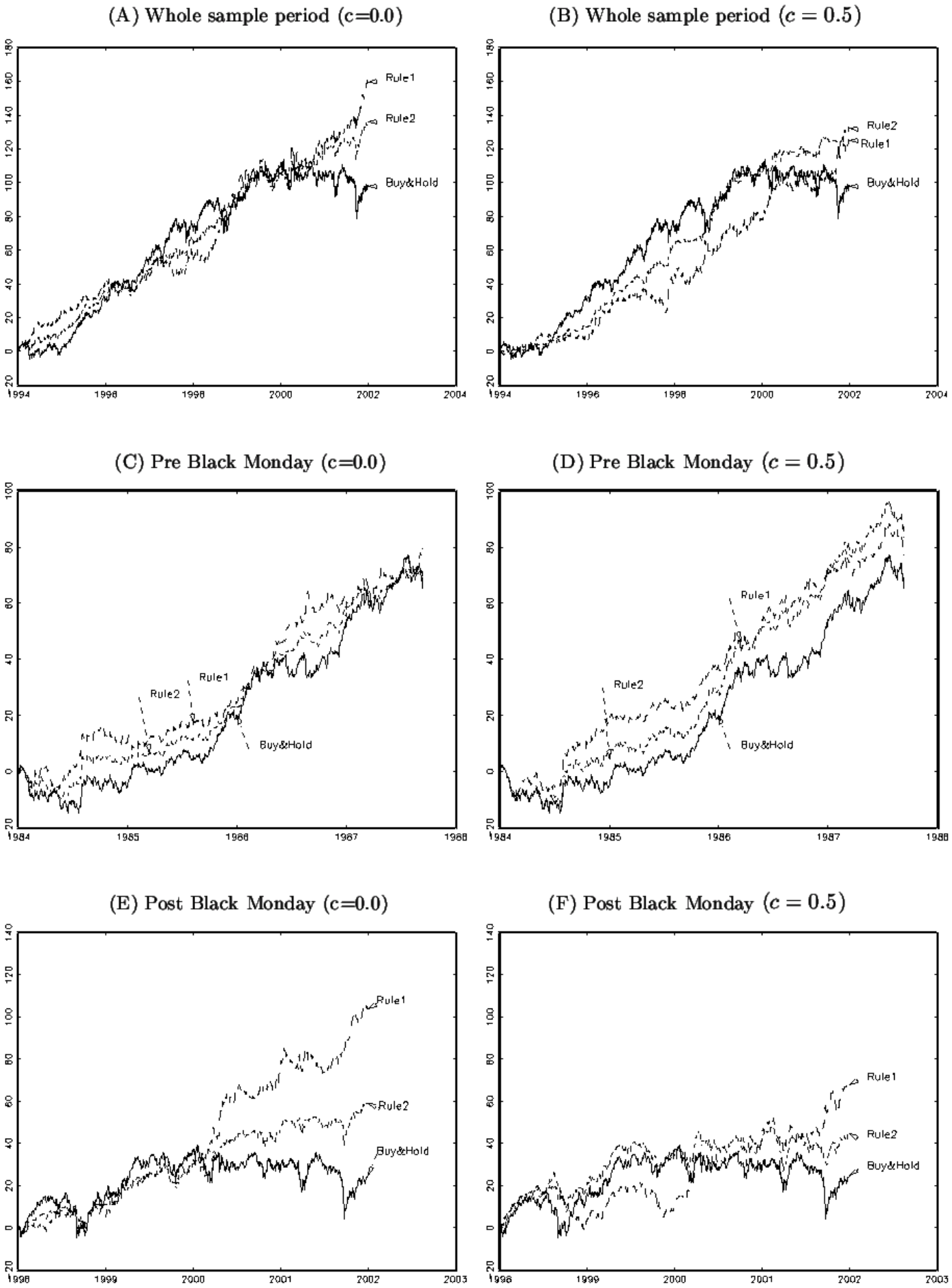

Figure 6: **Cumulative Daily Returns (%) of S&P500**

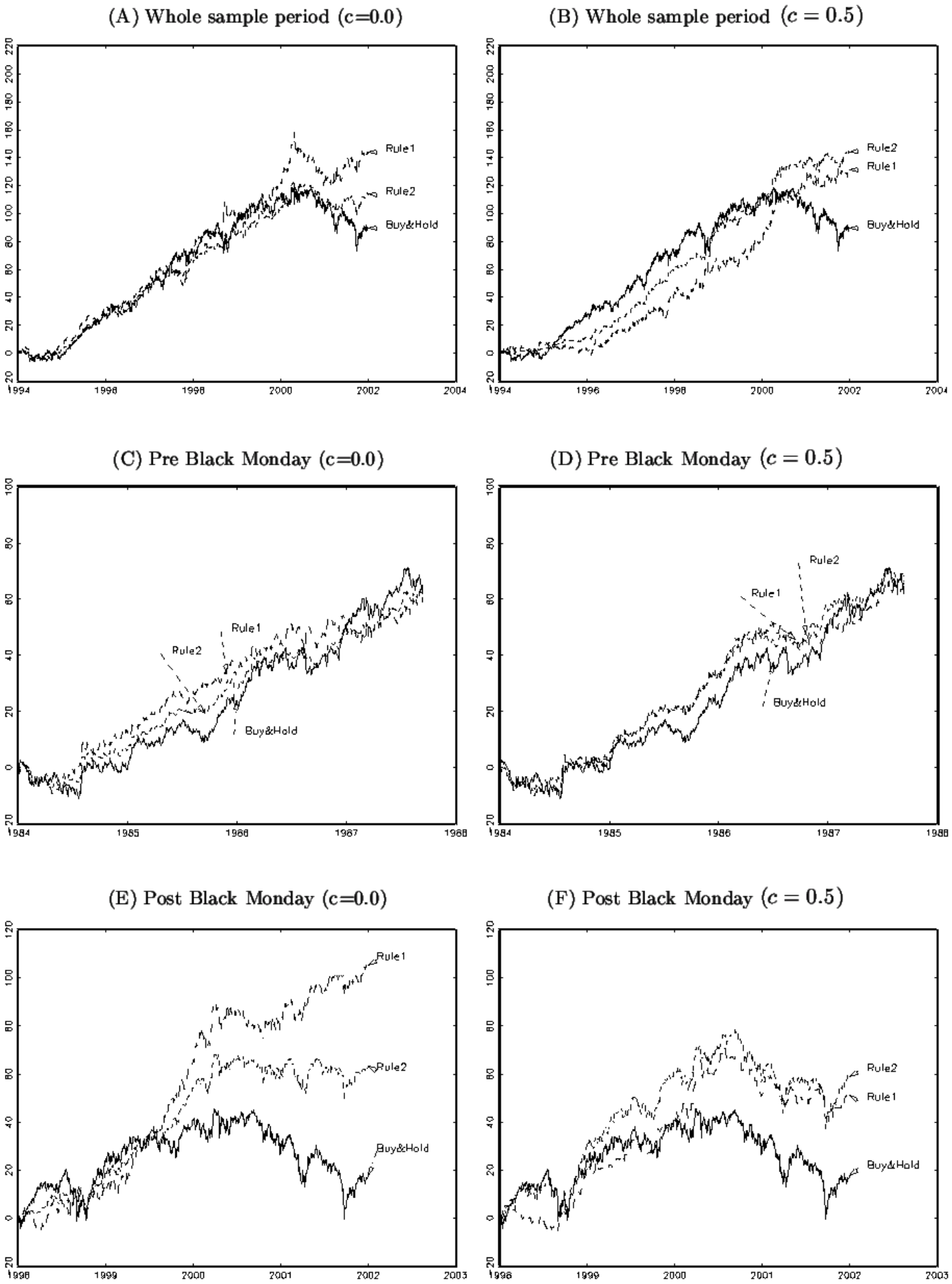

Figure 7: **Cumulative Daily Returns (%) of NYSE**

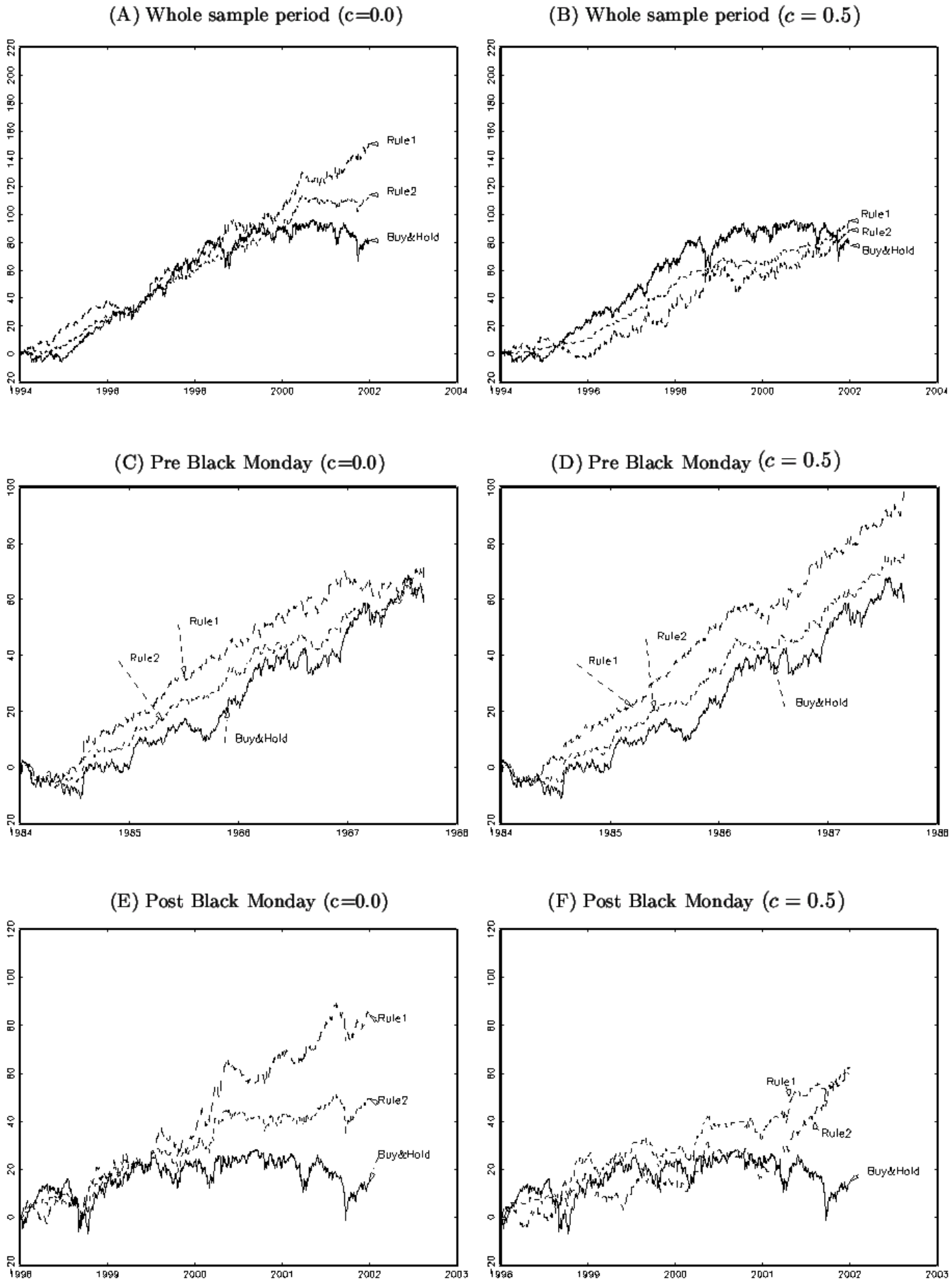

Figure 8: **Cumulative Daily Returns (%) of Nasdaq**