Market Timing of International Mutual Funds: A Decade after the Scandal

Vladimir Cherkassky, Fellow, IEEE and Sauptik Dhar.

***Abstract*— In early 2000's, the mutual funds industry was shaken by a series of scandals involving (legal) market timing and (illegal) late trading of mutual funds. These scandals have resulted in various regulations restricting investors' ability to exchange their mutual fund holdings. This paper applies data-analytic modeling techniques to test the effectiveness of market timing of international mutual funds. Our results indicate that whereas market timing strategies have been indeed profitable in the past, these strategies do not provide consistent improvement over traditional 'buy-and-hold' strategy during most recent period of 2009-2011. Ineffectiveness of market timing strategies seems to be directly related to increased volatility and changing statistical characteristics of the financial markets. These changes may warrant critical examination of trading restrictions placed on individual investors. These restrictions put individual investors at a disadvantage vs. institutional investors and hedge funds who invest in the same securities.**

***Index Terms*— Inefficient pricing, international mutual funds, interpretation of black-box models, market timing, predictive data analytics, market volatility, Vapnik-Chervonenkis (VC) theory.**

I. INTRODUCTION

According to conventional wisdom known as 'efficient market hypothesis' or EMH, market timing cannot be consistently profitable, see [1]. Yet it may be possible to profit consistently from inefficient pricing of certain assets, such as international mutual funds. Frequent trading, or timing of international mutual funds attempts to profit from daily price fluctuations, under the assumption that the next-day price changes may be statistically 'predictable' from today's market data [2-4]. Practical feasibility of market timing is the result of inefficient pricing of international funds by the mutual fund companies, so that the logical solution to this problem appears to be developing improved strategies for calculating daily Net Asset Value (NAV) of these funds [2,3]. However, the solution adopted by the Securities and Exchange Commission (SEC) and mutual fund industry was to place restrictions and fines on investors who engage in frequent trading activities; say hold their fund investments less than 60 days. Moreover, similar trading restrictions had been broadly placed on all domestic large-cap mutual funds as well, even though the pricing of such funds is known to be very efficient [2,3]; so it is not possible to profit from market timing of such funds. These trading restrictions create a dichotomy between small investors and sophisticated institutional players who can control their risk and market exposure on a minute-by-minute basis [5].

This paper pursues several objectives. First, we demonstrate application of predictive learning techniques for market timing of international mutual funds. This is, of course, a purely academic exercise, in view of current trading restrictions. However, it provides an interesting illustration of the predictive modeling methodology and several issues related to interpretation of black-box predictive models. Second, this exercise clearly demonstrates non-stationary (changing) nature of financial markets. Third, our modeling results raise doubts about the benefits/value of trading restrictions (reflecting past market conditions) in the current market environment.

Our study presents data analytic modeling using American Century International Growth Fund (symbol TWIEX) as a representative diversified international mutual fund. There is nothing special about this fund selection, and our results and findings generally hold for other diversified international funds. However, our findings will not hold for specialized international funds, e.g., Asian or developing markets funds. Market timing strategies are evaluated for two specific time periods:

– 2004-2005 period, as a representative period for past market conditions;

– 2009-2011 period representing current market conditions.

Again, there is nothing special about using the 2004-2005 period, and similar modeling results hold for other periods prior to 2008 (the beginning of current financial crisis).

The trading or market timing strategy for TWIEX generates a BUY or SELL signal at the end of each trading day, based on today's input indicators right before US stock market close at 4p.m. Eastern Standard Time. Effectively, by placing the BUY (or SELL) order today, you are betting that the price of this mutual fund will go UP (or DOWN) tomorrow, or the next trading day. Note that, the trading strategy uses two input indicators based on the daily closing prices of the following indices:

– SP 500 stock index (symbol ^GSPC);

– Euro-to-dollar exchange rate (symbol EURUSD=X).

Manuscript received November 15, 2011. This work was supported by NSF grant ECCS-0802056.

Vladimir Cherkassky is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis MN 55455 USA. (phone: 612 625-9597; fax: 612 625-4583; e-mail: cherk001@umn.edu).

Sauptik Dhar is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis MN 55455 USA. (e-mail: dharx007@umn.edu).

These two inputs should be good predictors for international mutual funds because (a) Asian and European markets close earlier than US market; (b) these funds are priced in US dollars.

According to this strategy, at any given day all funds are either fully invested in TWIEX or 100% in cash or other risk-free asset. This is a daily trading strategy, where at most one trading decision (transaction) can be initiated at any trading day. All modeling results assume no trading restrictions (on TWIEX) and zero transaction costs.

The trading strategy can be formalized as a binary classifier generating BUY or SELL decision rule based on the two inputs. The modeling methodology estimates this decision rule using training data (typically during one year period) and then applies it for trading during the next year (test period). The primary performance index of the trading strategy is the dollar gain (or loss) during the test period, vs. gain/loss of a simple buy-and-hold strategy for the same fund TWIEX. An additional performance metric is the market exposure, defined as the fraction of trading days an account is fully invested. Note that for a buy-and-hold strategy the market exposure is always 100%.

The rest of the paper is organized as follows. Section II presents detailed description of the predictive modeling approach and outlines several important modeling assumptions. Section III presents modeling results for 2004-2005 period. Section IV presents modeling results for current 2009-2011 period. Finally, summary and discussion are given in Section V.

II. Predictive Modeling Methodology and Assumptions

The trading strategy is implemented as a binary classifier using two inputs representing *daily percentage changes* of SP500 index (variable x_1) and euro-to-dollar exchange rate (variable x_2). For example, the first input x_1 is defined as x_1 = 100% * [SP500 (t) – SP500 (t-1)] / SP500 (t-1) where, SP500(t) is today's closing price, and SP500(t-1) is yesterday's (previous business day) closing price of SP500.

The binary output (or response) corresponds to BUY (+1) or SELL (-1) decisions for TWIEX. This classifier effectively tries to predict the next-day price movement of TWIEX based on today's change in SP500 and euro-to-dollar exchange rate. In this classifier model the inputs $\mathbf{x}$ correspond to *today's* daily percentage changes, whereas the output is *tomorrow's* change (Up or Down) of TWIEX net asset value. The classifier is estimated using available training data $(\mathbf{x}_i, y_i), i = 1, 2, ..., n$, where the class labels y_i corresponding to UP or DOWN daily changes of TWIEX are known. The trained classifier is then applied to new test data and its performance is evaluated as the total gain/loss during the test period.

This study uses two simple parameterizations for decision rule:

Linear: $$g(\mathbf{x}, \mathbf{w}) = w_1 x_1 + w_2 x_2 + w_0 \qquad (1a)$$

Quadratic:

$$g(\mathbf{x}, \mathbf{w}) = w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + w_0 \qquad (1b)$$

The decision rule for making predictions during test period is given by:

$$D(\mathbf{x}) = Sign(g(\mathbf{x}, \mathbf{w}^*)) \qquad (2)$$

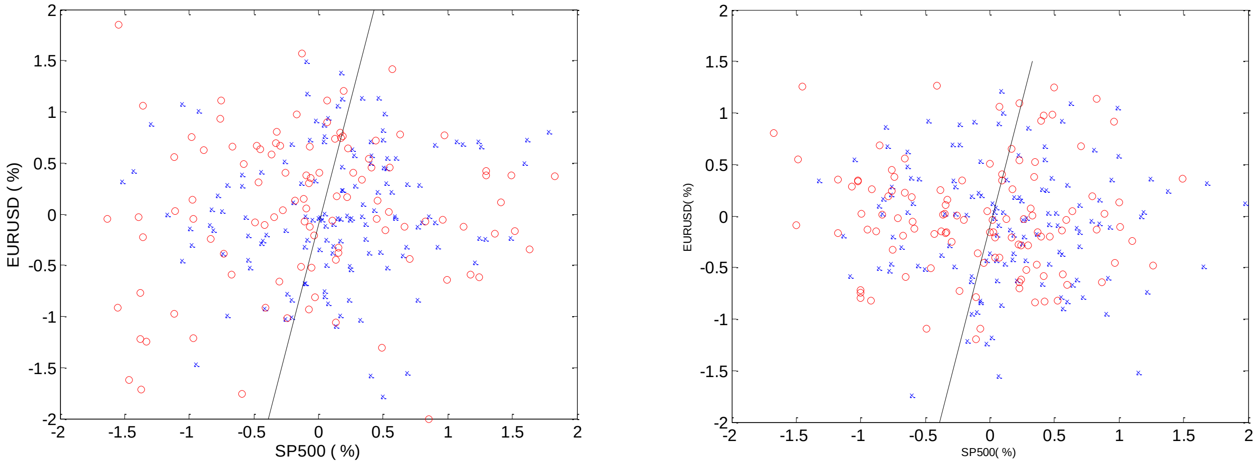
Here the model parameters (denoted as vector $\mathbf{w}^*$) are estimated by a learning method applied to training data. In particular, the linear model (1a) is estimated using Fisher Linear Discriminant Analysis (LDA) [6] and the quadratic model (1b) is estimated using a Quadratic Discriminant Analysis (QDA) [6] which applies the LDA method in an expanded five-dimensional feature space.

This modeling approach is based on several statistical assumptions. First, encoding of the input and output variables as daily percentage changes ensure that data samples $(\mathbf{x}_i, y_i)$ are approximately independent of each other. This i.i.d. property is a necessary condition for most practical learning algorithms. Another important assumption is that training and test data has similar (but unknown) statistical characteristics.

Predicting the next-day price movement (UP or DOWN) is impossible according to EMH, so consistently good predictive performance of estimated model can be used to disprove this hypothesis. There are two indices used for evaluating model's performance on test data:

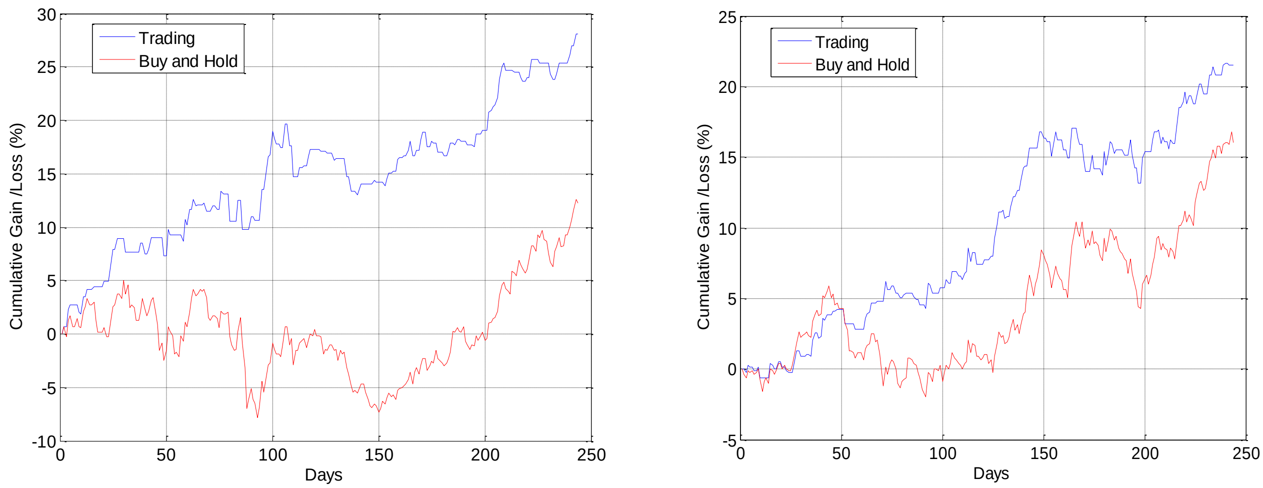
- total account gain (or loss) achieved by the trading strategy vs. total gain/loss of a simple buy-and-hold approach;
- total market exposure measured as a fraction (percentage) of days the trading account is fully invested in TWIEX.

Note that available data is very noisy, so minimizing the training error using an adaptive (flexible) classifier will likely result in overfitting and poor generalization. This motivates fixed low-complexity parameterizations, such as LDA (using 3 parameters) and a Quadratic Decision Boundary classifier (using 6 parameters) adopted in this study. So the number of parameters, or the VC-dimension, of such models is small relative to the number of training samples (~ 240 trading days). Then according to VC-theory [6, 7], if such a low-complexity model achieves good performance on the training data, it can be expected to generalize well for future (or test) data. Arguably, the quadratic parameterization (1b) is too complex for this data set, and it can potentially yield unstable models prone to overfitting. So this parameterization is used just to explore the possibility of useful *nonlinear* decision boundaries for this application.



(a) (b)

Fig.1. Fisher Linear Discriminant Model (LDA) along with training and test data (a) LDA with training data (year 2004). (b) LDA with test data (year 2005).



(a) (b)

Fig. 2. Effectiveness of the trading strategy using linear (LDA) model. (a) Account gain/loss during training period (year 2004). (b) Account gain/loss during test period (year 2005).

TABLE I
PEARSON CORRELATION COEFFICIENTS OF THE DAILY PERCENTAGE CHANGES OF TWIEX AND OTHER INTERNATIONAL MUTUAL FUNDS

Time Period	VGTSX	JSEAX	FIGRX
2004	0.9208	0.8820	0.9301
2005	0.7860	0.8744	0.8952
2010	0.9651	0.9551	0.9753
10/01/2009 – 09/30/2010	0.9689	0.9588	0.9744
10/01/2010–09/30/2011	0.9691	0.9663	0.9718

There are several training/test data set scenarios used in this study, including

– *Scenario 1*: year 2004 as training data and year 2005 as test data;
– *Scenario 2*: year 2004 as training data and year 2010 as test data;
– *Scenario 3*: training period ~ 10/01/2009 to 09/30/2010 and test period ~ 10/01/2010 to 09/30/2011.

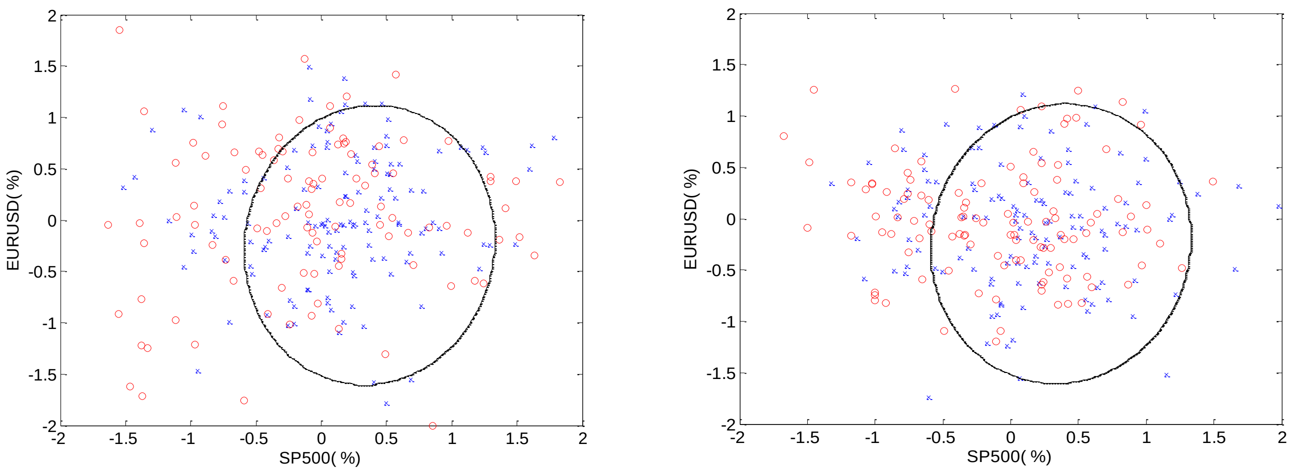
Here, Scenario 1 corresponds to 'past market conditions' and Scenario 3 corresponds to 'present market conditions'. Scenario 2 is used to illustrate performance of a trading strategy estimated using past market training data but applied under present market conditions. Under each scenario, we apply both methods (to estimate linear and quadratic decision boundary) using training data, and then evaluate an estimated trading strategy on the test period. These modeling results are presented in the next two sections. We have also performed additional comparisons using other learning methods (e.g., SVM, CART) for estimating the trading strategy and using different training/test periods. For example, it is possible to use year 2005 as training and year 2004 as test period, in order to gain confidence in the robustness of trading strategies. These additional results support our conclusions based on analysis of Scenarios 1-3, but they are not included due to space constraints. Moreover, there is nothing special about using the American Century International Growth Fund (TWIEX) as a representative diversified international mutual fund. Our results and findings will generally hold for other diversified international funds. This is evident from Table 1, showing

TABLE II
PERFORMANCE METRICS FOR TRADING STRATEGY USING LINEAR MODEL

	Error	Gain/Loss	Market Exposure
Training Period (2004)	41.15%	28.11%	51.85%
Test Period (2005)	43.62%	21.52%	57.61%

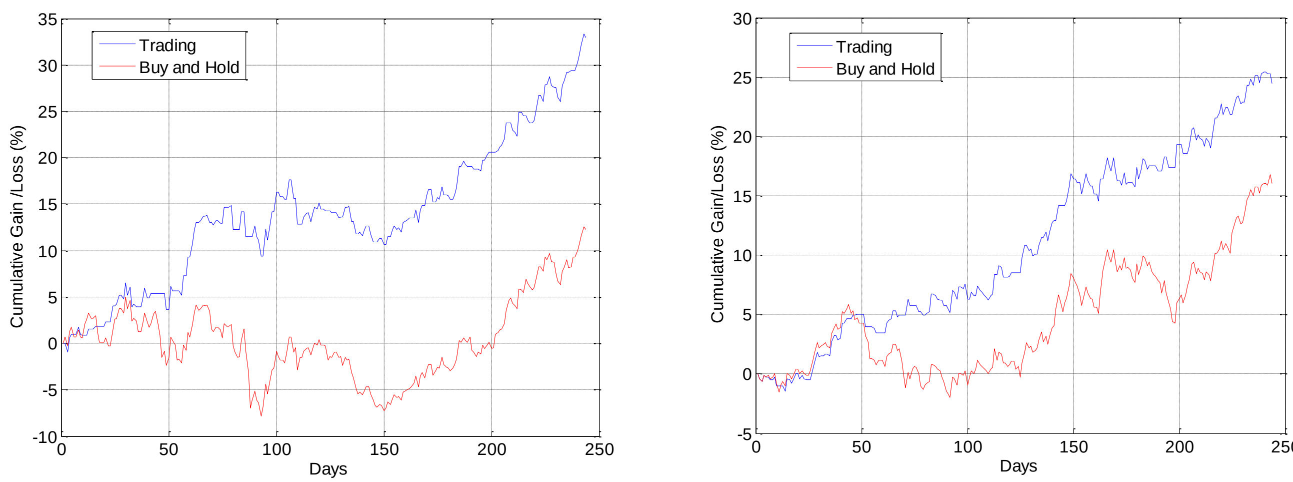
TABLE III
PERFORMANCE METRICS FOR TRADING STRATEGY USING QUADRATIC MODEL

	Error	Gain/Loss	Market Exposure
Training Period (2004)	45.27%	32.97%	67.08%
Test Period (2005)	39.09%	24.42%	69.96%



(a) (b)

Fig. 3. Quadratic Discriminant Analysis (QDA) along with training and test data. Area inside 'the circle' corresponds to a 'Buy' signal. (a) QDA with training data (year 2004). (b) QDA with test data (year 2005).



(a) (b)

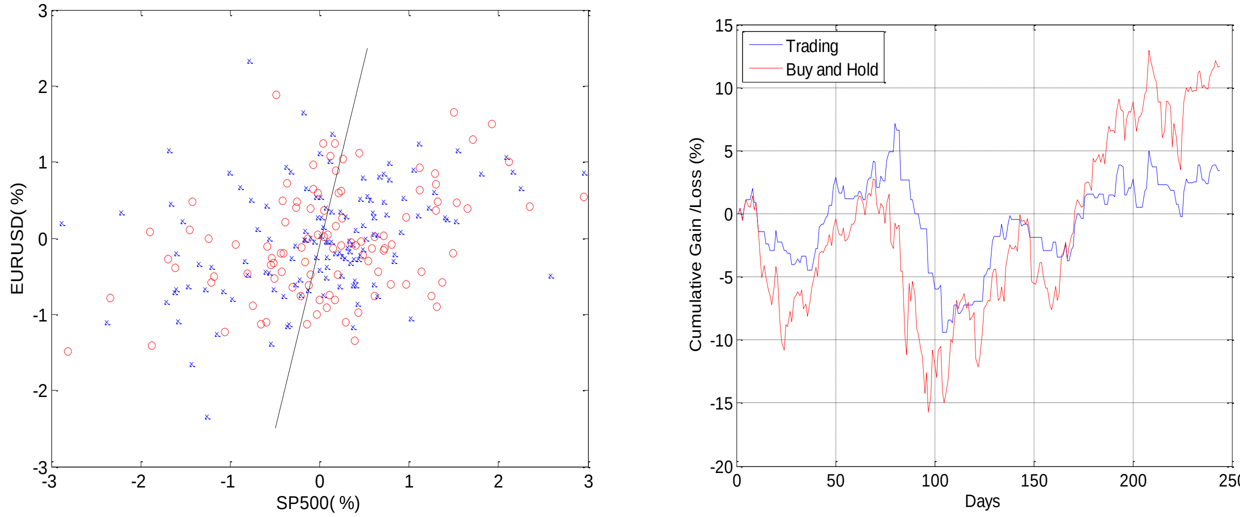
Fig. 4. Effectiveness of the trading strategy using quadratic (QDA) model. (a) Account gain/loss during training period (year 2004). (b) Account gain/loss during test period (year 2005).

(very high) correlation between the daily percentage changes of TWIEX and three other representative funds from different fund families: Fidelity International Discovery (FIGRX), Vanguard Total International Stock Index Fund (VGTSX), and JPMorgan International Equity A (JSEAX). Notice very high correlation observed for all annual time periods used for training and testing in this study.

III. MODELING RESULTS UNDER PAST MARKET CONDITIONS

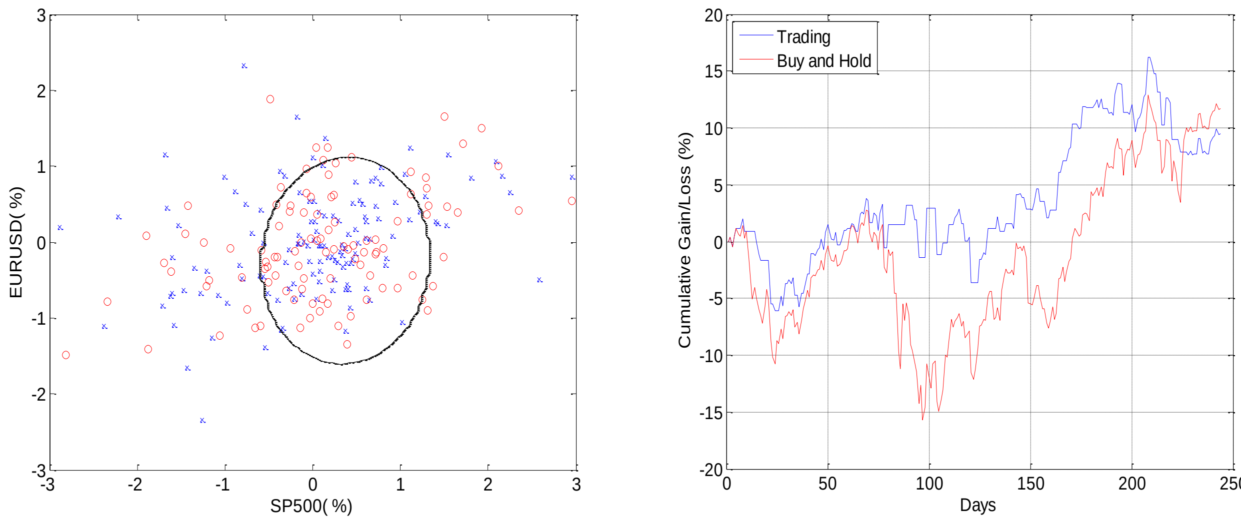
Under Scenario 1, the training data for Year 2004 is used to estimate the linear and the quadratic decision boundary. Fig. 1 shows the linear decision boundary along with the training data (2004) and test data (2005). The output labels (Up or Down) are indicated in blue and red color, respectively. Clearly, both the training and test data sets are very noisy and the two classes are heavily overlapping. Yet the estimated linear decision boundary yields a good trading strategy, as shown in Fig. 2. This figure shows the cumulative gain/loss of this trading strategy vs. the Buy-and Hold strategy for TWIEX.

Similarly, Fig. 3 shows the quadratic decision boundary and Fig.4 shows the performance of market timing based on this quadratic classifier. It is interesting to note that both models (linear and quadratic) have similar good performance



(a) (b)

Fig. 5. Effectiveness of the trading strategy using linear (LDA) Model (estimated using 2004 training data) during 2010 test period (a) LDA model with test data (year 2010) (b) Account gain/loss of the LDA model during test period (year 2010).



(a) (b)

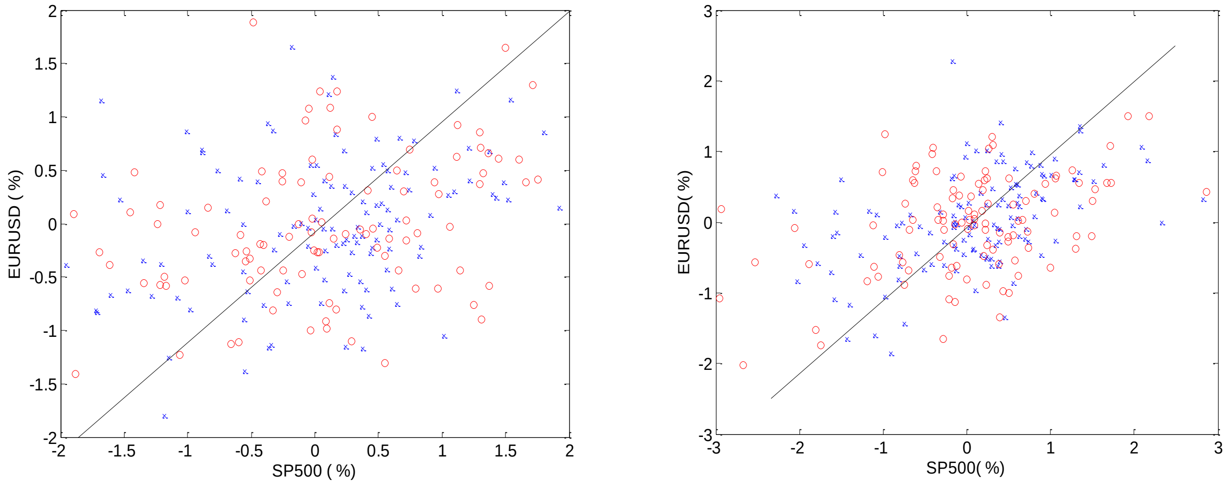
Fig.6. Effectiveness of the trading strategy using quadratic (QDA) model (estimated using 2004 training data) during 2010 test period. (a) QDA model with test data (year 2010) (b) Account gain/loss of the QDA model during test period (year 2010).

for test period (2005), even though they implement very different decision rules (see Fig. 1 vs. Fig. 3). Statistical performance indices for the linear and quadratic model are shown in Table II and III respectively. Note that the quadratic model has larger market exposure than the linear model (~ 70% vs. 50%). So the quadratic model is biased towards up-trending markets. This is evident from its superior performance during 2005 test period.

Next, we test the performance of trading strategy (estimated using 2004 training data) using year 2010 test period. These results are shown for the Linear Discriminant Analysis (LDA) model and for the Quadratic Discriminant Analysis (QDA) model, in Figures 5 and 6, respectively. Under this setting, the performance of these trading strategies is no better (or worse) than simple Buy-and-Hold approach. Clearly, these trading strategies (estimated using 2004 market data) fail to capture statistical characteristics of present market conditions. An obvious way to address this issue is to train the model using present market data. This approach is investigated in the next section.

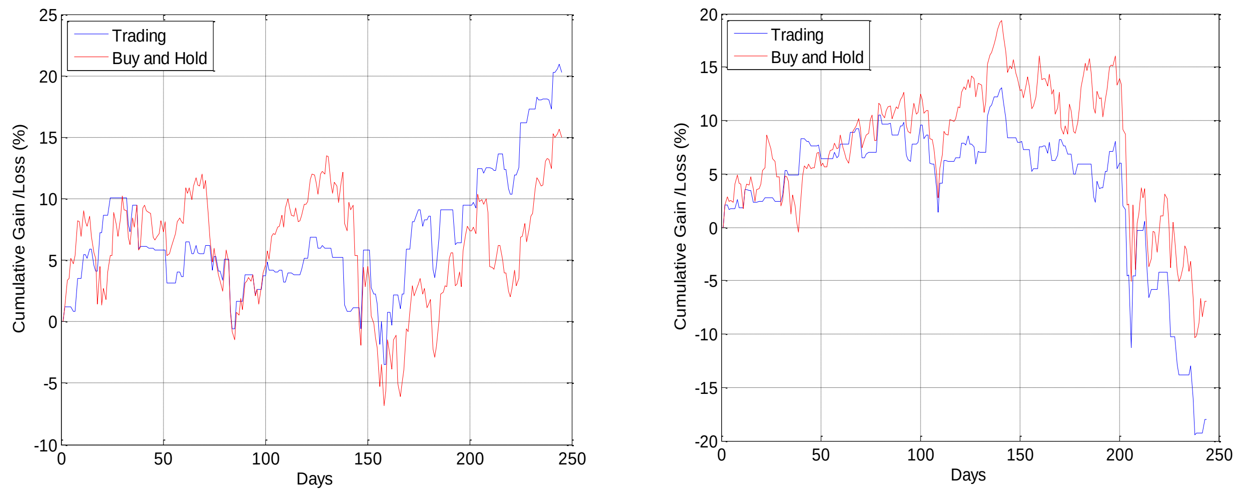
IV. Modeling Results under Present Market Conditions

This section investigates performance of trading strategies under present market conditions, using 2009-2011 market data for both training and test periods. Under this setting (Scenario 3), the training data corresponds to the period from Oct. 1, 2009 to Sep 30, 2010, and test period is from Oct. 1, 2010 to Sep 30, 2011. As before, the training data is used to estimate the linear and the quadratic decision boundary. Fig. 7 shows the LDA decision boundary along with the training and test data. The output labels (Up or



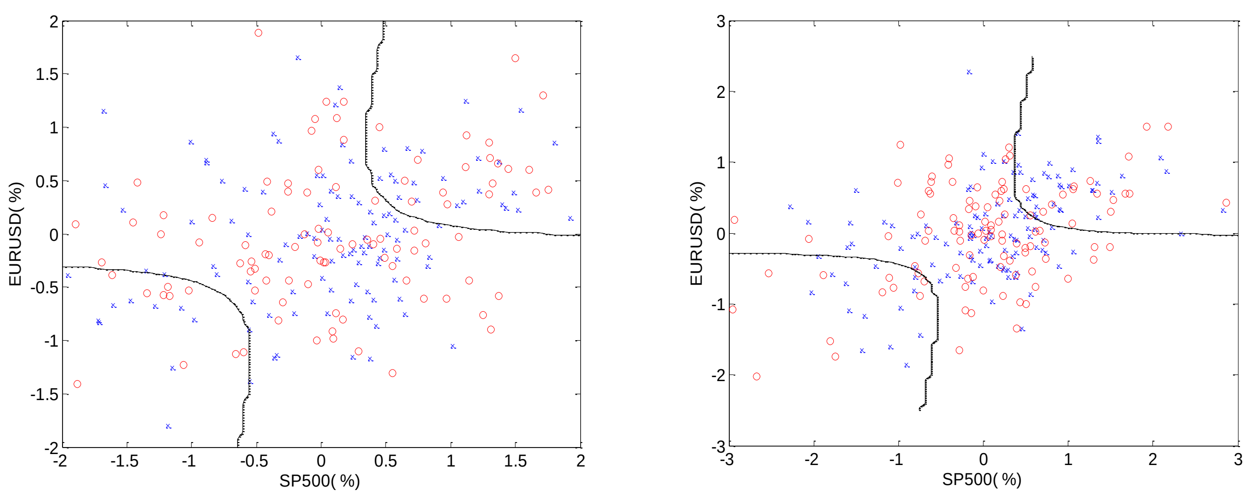
(a) (b)

Fig. 7. Fisher Linear Discriminant Model (LDA) along with training and test data (a) LDA model with the training data (year 2009-10) gives a training error of 51.23%.(b) LDA model with the test data (2010-2011) gives a test error of 51.44%.



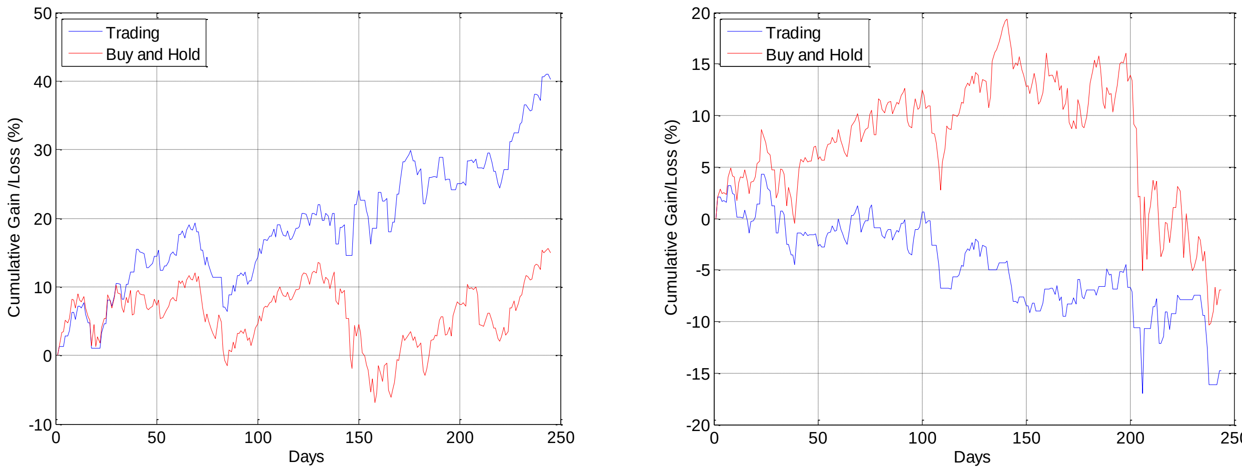
(a) (b)

Fig. 8. Effectiveness of the trading strategy using linear (LDA) model. (a) Account gain/loss during training period (year 2009-2010) with a market exposure of 45.49%. (b) Account gain/loss during test period (2010-2011) with a market exposure of 48.56%.



(a) (b)

Fig. 9. Quadratic Discriminant Analysis (QDA) Model along with training and test data. (a) QDA Model with the training data (year 2009-10) gives a training error of 44.26%.(b) QDA Model with test data (2010-2011) gives a test error of 53.09%.



(a) (b)

Fig. 10. Effectiveness of the trading strategy using quadratic (QDA) model(a) Account gain/loss during training period (year 2009-2010) with a market exposure of 67.21%.(b) Account gain/loss during test period (2010-2011) with a market exposure of 65.02%.

Down) are indicated in blue and red color, respectively. Fig. 8 shows the cumulative gain/loss of this trading strategy vs. the Buy-and-Hold strategy for TWIEX. Modeling results for the quadratic (QDA) classifier are shown in a similar format in Figs. 9 and 10.

These results indicate that market timing does not work well under current market conditions, i.e. it does not provide any improvement over traditional Buy-and-Hold approach. Moreover, the quadratic model appears to have serious overfitting problem, as evident from its performance on the training and test data (see Fig. 10).

V. SUMMARY

In conclusion, we present critical discussion of the modeling results in Sections III and IV. Our discussion is focused on:

1) interpretation of black-box data-analytic models under VC-theoretical/ predictive modeling methodology [4,6].
2) further statistical analysis nature of financial markets and understanding factors responsible for poor performance of market timing under current market conditions.
3) several policy implications regarding current restrictions on frequent trading of mutual funds.

Interpretation of predictive models: Modeling results for the 2004-2005 period presented in Section III illustrate application of two predictive models using linear and quadratic parameterization. These results indicate that *both models* yield very good performance, i.e. consistently outperform the Buy-and-Hold strategy. Both trading strategies also result in a low market exposure, as the trading account is out of the market (in cash) about 40% of the time.

Classical statistical interpretation of data-analytic models assumes that such models provide close approximation to the 'true' model or the best possible model that can be estimated using (asymptotically) large number of samples. In contrast, under predictive modeling methodology the notion of a single true model becomes problematic [6, 8]. So the usual question (posed by classical statisticians) which model most accurately describes the training data is difficult to resolve [8]. In fact, the two data-analytic models, linear and quadratic, analyzed in Section II reflect two *different* successful trading strategies. These strategies reflect *different* properties of the data which yield good prediction performance. These strategies can be explained by financial experts and traders. However, understanding and explanation of these models requires application domain knowledge, and *cannot rely only* on a data-analytic model. In particular, the linear decision boundary in Fig. 1 can be interpreted as the rule: '*Buy* if SP500 is up today, otherwise *sell*'. This rule has simple causal interpretation: the next-day direction of foreign markets tends to follow today's change of the US stock market. The quadratic decision boundary shown in Fig. 3 has a different interpretation: '*Buy* if today's change in SP500 is not_large AND the change in the EURO exchange rate is not_large, otherwise *sell*'. This rule is more complex, but also has a common-sense explanation: *very large changes* of the input variables usually occur in response to the news (such as earnings reports, economic statistics) released *in the morning* when European markets are still open. Hence, such information is likely to be already reflected in the closing prices of European equities. Note that both interpretations are based on understanding of the problem at hand, i.e. knowledge about financial markets, opening/closing time of the American and European markets etc. This knowledge *cannot* be derived from black-box predictive models alone.

Changing statistical nature of financial markets: Modeling results in Section IV demonstrate ineffectiveness of market timing of international mutual funds under present market condition. The reason is that statistical characteristics of the stock market have changed dramatically. For example, for 2004-2005 data there was no correlation between the two input variables in the trading model (daily fluctuations of the SP500 index and the euro-to-dollar exchange rate), whereas for 2009-2011 period this correlation is quite large (~ 0.4). Likewise, during 2009-2011 period, European markets exhibit very high correlation with US stock market. Namely, during 2009-2011 the correlation between daily fluctuations of SP500 and TWIEX is ~ 0.9, whereas during 2004-2005 period this correlation is much lower, in the range [0.4 - 0.6]. There may be several reasonable explanations for these changing market characteristics:

- international markets are more tightly linked (correlated) to US stock market, due to globalization and electronic trading.
- during 2009 – 2011 period, it is the US stock market that follows European markets (and not vice versa). This seems to be well supported by a large correlation between daily fluctuations of SP500 and the daily closing prices of the international mutual fund TWIEX.
- the procedure for calculating the daily NAV value of TWIEX has changed, in order to reflect more accurately the daily changes of the US stock market.

Each of these explanations or their combination may be true. However, it is not possible to derive true 'causal' explanation based on the data-analytic model alone.

Policy Implications: Finally, we comment on possible implications of this study with regard to current restrictions on the frequency of mutual funds redemptions. Currently, all international (and most domestic) mutual funds impose restrictions on frequent fund exchanges. Frequent fund exchanges are not illegal, but are considered to be 'unethical' by the mutual fund industry. There are two main reasons for these restrictions:

1) investors engaged in market timing consistently achieve gains, at the expense of other long-term mutual fund holders;
2) frequent trading disrupts investment strategies of the fund managers and results in extra administrative and trading costs.

However, from the investors' perspective, these restrictions prevent individual investors from managing risk and market exposure. This risk management becomes especially relevant under present market conditions, when investors are faced with negligible / negative 'long-term' market returns over the past decade, accompanied by increased short-term market volatility. Whereas many professional investors and hedge funds can manage risk in real-time via various sophisticated instruments, most individual investors in mutual funds can control risk only via asset allocation and buy-and-hold approach.

The results of our study suggest that under current market conditions, the restrictions on market timing of mutual fund do not make sense. In fact, as shown in Section IV, application of both market timing strategies yields inferior performance (vs. Buy-and-Hold) during test period. So an investor consistently applying market timing will lose money, and this will effectively result in *improved performance* for the majority of 'buy-and-hold' investors. Based on this reasoning, it appears that trading restrictions (introduced in early 2000's) need to be critically re-examined to reflect present market conditions. Currently, these restrictions appear to serve mainly the interests of the mutual funds industry rather than individual investors.

References

[1] "Market Timing", [Online], Available: http://en.wikipedia.org/wiki/Market_timing. [Accessed: Nov 14, 2011].

[2] E. Zitzewitz, "Who cares about shareholders? Arbitrage proofing mutual funds", *Journal of Law, Economics and Organization*, vol .19 (4), pp. 245-280, 2003.

[3] T. Frankel, and L.A. Cunningham, "The mysterious ways of mutual funds: market timing", *Annual Review of Banking and Financial Law*, vol. 25, pp. 235-293, 2006.

[4] V. Cherkassky, F. Mulier, and A. Sheng, "Funds exchange: and approach to risk and portfolio management" , *Proc. Int'l Conf. on Computational Intelligence in Financial Engineering (CIFEr)*, New York, 2000,pp. 3-7.

[5] H. Gao, and V. Cherkassky, "Real-time pricing of mutual funds", *IJCNN*, Vancouver, 2006, pp. 2402 – 2408.

[6] V. Cherkassky, and F. Mulier, *Learning from Data Concepts: Theory and Methods*, 2nd ed. NY: Wiley, 2007.

[7] V. N. Vapnik, *Estimation of Dependencies Based on Empirical Data. Empirical Inference Science*: Afterword of 2006. New York: Springer, 2006.

[8] L. Breiman, "Statistical Modeling: the Two Cultures", *Statistical Science*, vol. 16(3), pp. 199-231, 2001.