Estimating NAV of Domestic Mutual Funds in Real-Time

Patrick Delaney[a]

[a] University of Minnesota, Undergraduate, Institute of Technology, Electrical Engineering

ABSTRACT

This paper investigates the use of regression modeling techniques to predict the Net Asset Value of Domestic Mutual Funds in Real Time. We discuss how this ability to predict may be used in a certain trading strategy, and what this prediction methodology possibly implicates for current frequent-trading restrictions in the mutual fund industry.

Received 3 May 2007.

INTRODUCTION

What is a Mutual Fund?

A mutual fund is a form of collective investment that pools money from many investors and invests their money in stocks, bonds, short-term money market instruments, and/or other securities. In a mutual fund, the fund manager trades the fund's underlying securities, realizing capital gains or losses, and collects the dividend or interest income. The investment proceeds are then passed along to the individual investors. The value of a share of the mutual fund, known as the net asset value per share (NAV), is calculated daily based on the total daily value of the fund shares (τ = i:k) divided by the number of shares currently issued and outstanding (n), adding cash and equivalent holdings (X) and subtracting liabilities (L).

$$NAV \cong \frac{\left[\sum_{i=1}^{k} \tau(i) + X - L \right]}{n}$$

This NAV is calculated once per day at the end of each trading day. For mutual funds trading on the New York Stock Exchange, this occurs between 4:00 and 5:00 pm eastern time. Our goal is to be able to predict the NAV of a given domestic mutual fund given certain "index inputs," which we shall discuss later.

Motivation - Mutual Fund Brief History

We have selected to investigate mutual funds in the United States for this paper. Mutual funds are relatively new in the United States – their inception was in 1924, when the "Massachusetts Investors Trust" began. Today, the mutual fund industry includes over 8,000 mutual funds, pursuing a laundry list of investment objectives. Since 1980, the popularity of mutual funds has skyrocketed. Ownership changed from 6% of all households in 1980 to 27% in 1993 [1]. This means that industry assets have increased as well – growing from $293 billion in 1980 and surpassing $2 trillion in 1993. As of April 2006, there were 8,606 mutual funds that belong to the Investment Company Institute (ICI), the national association of investment companies in the United States, with combined assets of $9.207 trillion [2].

In September 2003, the United States mutual fund industry was beset by a scandal in which major fund companies permitted and facilitated "late trading" and "market timing." These companies used NAV calculation policies which did not account for certain inefficiencies in the market due to worldwide timing effects – allowing market timers to get a much higher return on investment than investors who "buy and hold." This anomaly was said to dilute the market and hurt long term shareholders [3]. As a result, many mutual funds adapted their NAV calculation policies and trading policies. Mutual funds now often require a 60-day waiting period after the first purchase of the mutual fund before choosing to sell it. One can see these policies outlined in the prospectus of any mutual fund. These policies have almost universally been applied to all mutual funds across the country – whether they be domestic or international in nature. Given this change, the question now arises, "why apply these frequent-trading policies across all mutual funds?" Is this restriction really in the best interests of mutual fund investors? If the assumption that, "frequent trading will affect the NAV of a mutual fund," is wrong – then wouldn't arguably the new policies hurt mutual fund customers rather than help them by restricting their ability to trade when they chose to?

OBJECTIVES & BACKGROUND

Our Objective

Our objective in this project is to *prove that domestic mutual funds NAV prices can be predicted by domestic market indices using "predictive modeling."* "Predictive modeling" means applying some form of mathematical model from one time interval to another. Specifically, we base some form of model on an earlier time interval and applied to a latter. The assumption being made is that the future will be "stochastic" or somehow similar to the past. There a variety of models which can be used – in this paper we base our model used off of the best model found in, "Real-Time Pricing of Mutual Funds" by Hui Gao and Vladimir Cherkassky [4].

Assumptions: The Mutual Fund Problem Domain

In essence, mutual funds are driven by their ability to realize a gain. All mutual funds must realize some sort of gain comparable to the US financial markets. As such, mutual funds use a variety of strategies to set benchmarks for their gains. Mutual fund managers attempt to beat or at the very least mimic these benchmarks by trading similar funds and paying close attention to the market to seize opportunities for increased gain. These benchmarks and trading strategies by law must be listed in the mutual fund prospectuses. Mutual funds will almost always benchmark themselves against a stock market index such as the S&P500, the DJIA or the Semiconductor Index, or some combination of indices. These benchmarks typically do not change over time – a given mutual fund will follow its benchmark for its entire lifetime. At this point it should be noted that the benchmarks which mutual funds follow are traded and priced in real-time. Indices such as the S&P500 and the DJIA move throughout a given day and are priced every few seconds – while as mentioned above mutual funds are priced only once per day at the end of the day. Using this fact we can presume that the price of a mutual fund on a given day will depend upon what happened to its associated index or indices in real-time throughout the day. Hence using this principle and "predictive modeling" we can predict the NAV of mutual funds.

EXPERIMENTAL SETUP SOFTWARE

Matlab

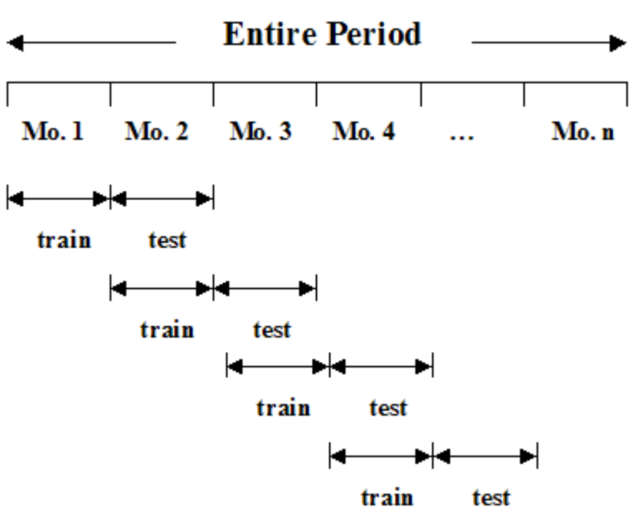
The software we used was MATLAB, which is a numerical computing environment and programming language. MATLAB allows easy matrix manipulation and various forms of plotting of functions and data. It specializes in numerical computing, which was ideal for this application.

EXPERIMENTAL DESIGN

Assumptions: Training and Test Data Used With Mutual Funds

Setting the Sample Space Size

There is no "one fits all" predictive model for mutual funds, all funds have different objectives. Neither is there any "one fits all" solution for a given mutual fund over time. Training and testing periods may range anywhere from one month to six months. This is because the training sample space must be sufficiently large to obtain an accurate model - and yet at the same time a sample training space cannot be too large as behavior may not remain stochastic. Hence we must try to estimate a "happy medium" range that applies the training period accurately to a given range. After trial and error we found that the best training and testing combination was to use "one month training setup" and "one month testing." We found that "one month" was optimally defined as 20 scrolling trading days (as shown in the figure below).



Input Variables: Exchange Traded Funds

Our model utilizes input variables which directly mimic financial market indices. These input variables are known as Exchange Traded Funds (ETFs). For the purposes of this paper, it can be accepted that ETFs are equivalent to market indices. We initially selected the ETFs based on their described similarity to the mutual funds themselves. For example, if the mutual fund was described as a "large cap growth fund" we would attempt to model it using a number of "large cap growth fund" ETFs. A general description of mutual ETFs used can be found in the appendix.

Mutual Funds Used

Fidelity Magellan [FMAGX]

The Fidelity Magellan Fund is described as a "large cap growth fund." Viewing the prospectus, the purpose of this fund is to seek growth through domestic, foreign and multinational issuers. As expected this fund prohibits frequent trading. In addition, the fund's pricing is noted to be priced based upon market conditions, interest rates, and in response to other economic, political or financial developments. It does invest in foreign issuers, but no more than 40% of the fund's assets may be invested in companies operating exclusively in any one foreign country [5]. Based on the description in the prospectus, we expect this mutual fund to correlate to a large-growth index such as the S&P500, DJIA or NASDAQ.

Fidelity OTC [FOCPX]

The investment seeks capital appreciation. The fund normally invests at least 80% of assets in securities principally traded on NASDAQ or an OTC market. Securities that begin to trade principally on an exchange after purchase continue to be considered eligible securities for purposes of the 80% policy. It may also invest the fund's assets in non-OTC securities. The fund will invest more than 25% of the fund's total assets in the technology sector. The fund is non-diversified [6]. Given this description we expect Fidelity OTC to correlate to the NASDAQ ETF (QQQQ) and perhaps other technology indicators such as the Semiconductor Index (SME).

Fidelity Contrafund [FCNTX]

The investment seeks capital appreciation. The fund normally invests primarily in common stocks. It may invest in securities of companies whose value is not fully recognized by the public. The fund invests in both domestic and foreign issuers. It may invest in "growth" stocks or "value" stocks or both. The advisor uses fundamental analysis of each issuer's financial condition and industry position and market and economic conditions to select investments [7]. Since this stock trades "may invest in securities not recognized to the public," this may indicate that this mutual fund follows the

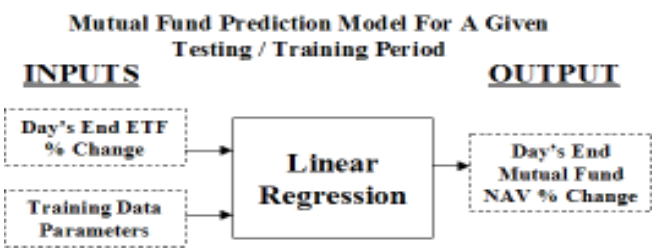
NASDAQ more than other indices – due to the NASDAQ's tendency to follow higher-growth stocks.

American Funds [AMCPX]

The investment seeks long-term growth of capital. The fund invests primarily in stocks of issuers located in the U.S. but may invest in non-U.S. securities to a limited extent. It generally purchases growth-oriented, equity-type securities which involves large price swings and potential for loss. The management focuses primarily on companies with attributes that are associated with long-term growth [8]. Similar to FMAGX, we expect this mutual fund to correlate to a large-growth index such as the S&P500, DJIA or NASDAQ.

Regression Formulation

The specific mathematical model we used was "Linear Regression." This was found to be the optimal method in the paper by Gau and Cherkassky. Hence our model has the form shown in the figure.



Quality Control

We evaluate the accuracy of our models by using three principle methods. The first method is viewing the box plots of the errors over time. The errors were calculated by percent error and stored in a vector. Using this vector we were able to construct box plots which give us a descriptive statistical view of how accurate the synthetic model was compared to the testing period. Secondly, we can look at a graph of the synthetic prices compared to the actual prices. This comparison graph is essentially a graph of our predicted synthetic fund on the same time scale as the

TESTING & RESULTS

Three-Year FMAGX, FCNTX, FOCPX and AMCPX Results
All four mutual funds have similar results in terms of their performance. Given the proper ETF input selection, the box plots of each were shown to be relatively stable over time. Errors normally stayed within the 2% to 4% range for every given month. Also upon viewing the comparison graphs we see that there is a strong visual correlation. Both of these displays can be seen for each mutual fund in the appendix.

Further, we can see the weight of each coefficient by looking at their averages and standard deviations. First, we look at the FMAGX. Since it is a large cap growth fund, we expect that it would be correlated to the SP500, DJIA or NASDAQ. Upon testing with our regression model, we find that this was true – the best indices tested was the S&P500 ETF (SPY) and the DJIA ETF (DIA). The results of this can be seen in the appendix. Looking at the mean and standard deviations below we can view that the SPY had the greatest impact:

FMAGX, 1-Month Train Coefficients 20 April 04 through 29 March 2007			
Coefficient	**w0**	**SP500**	**DJIA**
Avg.	0.00	1.27	-0.04
Sdv.	0.00	0.57	0.47

We found that FCNTX, despite being described as being "technology-based" has a stronger correlation to the energy sector index ETF (XLE) than the semiconductor index (SMH). The standard deviation of XLE was much less than that of SMH when modeled along with QQQQ. The final results of the coefficient mean and standard deviation can be seen below.

FCNTX, 1-Month Train Coefficients 20 April 04 through 29 March 2007			
Coefficient	**w0**	**XLE**	**QQQQ**
Avg.	0.00	0.16	0.53
Sdv.	0.00	0.07	0.11

For FOCPX, the without a doubt the strongest correlation occurred with the NASDAQ, as expected. This can be seen in the results shown below, which indicate a strong mean and low standard deviation of linear regression connection to QQQQ.

FOCPX, 1-Month Train Coefficients 20 April 04 through 29 March 2007				
Coefficient	**w0**	**SP500**	**QQQQ**	**SMH**
Avg.	0.00	0.08	0.94	0.057
Sdv.	0.00	0.11	0.11	0.091

AMCPX, 1-Month Train Coefficients 20 April 04 through 29 March 2007			
Coefficient	**w0**	**SP500**	**QQQQ**
Avg.	0.00	0.007143	0.580517
Sdv.	0.00	0.071697	0.089823

DISCUSSION

Predictive Modeling Accuracy

It is apparent that mutual funds can be predicted to some degree based on the results of this study. But what exactly might this level of prediction indicate in a practical sense? Before we make final conclusions, we will discuss an assumption of how the efficient market hypothesis may apply, and how one might be able to build a trading strategy base on certain market efficiencies from the new knowledge we have created.

Assumptions: How Efficient Market Hypothesis Applies

Before attempting to discuss the results of our NAV prediction via market indices, we should first consider a widely held hypothesis known as the Efficient Market Hypothesis (EMH). This theory, developed by Professor Eugene Fama at the University of Chicago Graduate School of Business in the early 1960s, states that it is not possible to consistently predict (and hence outperform) a highly efficient financial market — appropriately adjusted for risk — by using any information that the market already knows, except through luck. This essentially means that all decision-making information can normally be assumed to be available to everyone participating in the market, causing the attempts to outperform predict (and thus outperform) difficult [9]. This theory supposedly arises from the notion that there are so many intelligent decision-makers in the marketplace – that the markets will "self-correct" given any new information.

Given the sheer volume of publicly available price data and widely available computing power today, this hypothesis remains an important in considering practical or theoretical implications of our results. Currently, as discussed above, mutual funds policies are in place which prevent investors from trading frequently. If mutual fund investors were allowed to trade frequently, they would no doubt be attempting to use the same computing power and public data that we have used for this investigation. Hence the NAV fluctuation patterns of those funds would undoubtedly differ as investors respond to market information. The question remains: prove that domestic mutual funds NAV prices can be predicted by domestic market indices using "predictive modeling

However, the mutual fund scandals of 2003 involved just that: market inefficiencies. Market timers' approach was to use empirical data that exists in the stock market and apply their own form of analysis to create new information that the market does not know. Hence successful market timing does not violate EMH and market timers were able to receive 35-70% returns per year [10]. In our case, we are utilizing "data" that is accessible to all players in the marketplace – but we are creating new "information" or "knowledge" that not all players have access to. Since they do not have access to this information, it creates a market inefficiency.

Building A Trading Strategy

One can theoretically "buy low and sell high," using the right market predictions. The idea here is that when the mutual fund's NAV is lower than what an accurate synthetic model shows that it "should be," we would want to buy the mutual fund. On the other hand, if it is shown to be higher than what the synthetic model shows that it "should be" we should sell the mutual fund. This is shown illustrated in the appendix under "building a trading strategy." To do this, we look at the model and the actual price over the past "n" days and average the differences. An equation form of this can be seen also in the appendix in the same section. Then using this factor, "delta_n" we compare it to our decision boundry "b". We set b at a fixed percent or value – in our AMCPX case the optimal value found was b = +/-0.06% of the difference between our synthetic and actual NAV (synthetic-actual). So when the average difference over the past 3 days is 0.06% of both NAVs, this means that the mutual fund is under-priced, so a BUY signal flags. However if the opposite boundary of -0.06% is reached, the mutual fund is over-priced and a SELL signal flags. Using this model and a 3-day waiting period for buy/sell we receive the performance shown in the appendix, which indicated a roughly 8% improvement change from what we would have seen with a buy and hold model. We show the results where the green is our asset value if we start out at $30,000 and follow the program's instructions and blue is if we start out at $30,000 and hold. For comparison, in the appendix is also shown the identical trading strategy with a less accurate model. In this model we used ETFs SPY and DIA as inputs. One can easily observe that the return patterns are similar – but not as efficient as with the former inputs. This is because the model is not as accurate, as can be observed by the monthly box plots shown in the appendix. This clearly shows the importance of an accurate trading strategy – and the hence the principle that greater levels of accuracy lead to greater market advantages.

CONCLUSION

First, we have proven that domestic mutual funds NAV prices can be predicted by domestic market indices using predictive modeling. This is done using the principle of creating new knowledge using publicly available data and computational methods. The further question that remains is – does the restriction of mutual fund trading hurt or help other mutual fund investors? The answer to this is highly speculative. If the restrictions did not exist, then it is reasonable to assume that many may attempt to use the trading strategies outlined in this paper. This would in affect change the NAV of a mutual fund, since withdrawing and depositing cash into the mutual fund has an affect on the price. With a few investors behaving in this manner, this would be barely noticeable. But with many investors acting this way – the NAV may fluctuate significantly due to frequent trading. So since it is beyond the scope of this paper to determine whether a large number of investors would behave in this manner, it is too difficult to predict how the NAV would be affected. However, given the principle that "more accurate synthetic models lead to more profitable trading strategies" – it is certain that the models outlined in this paper may have to adapt if a large number of investors were engaging in this behavior. Finally, since the trade restrictions normally prohibit only fast selling of the mutual funds after 60 days – then it may be possible to look for buying and selling signals outside of these 60 day windows of time.

APPENDIX

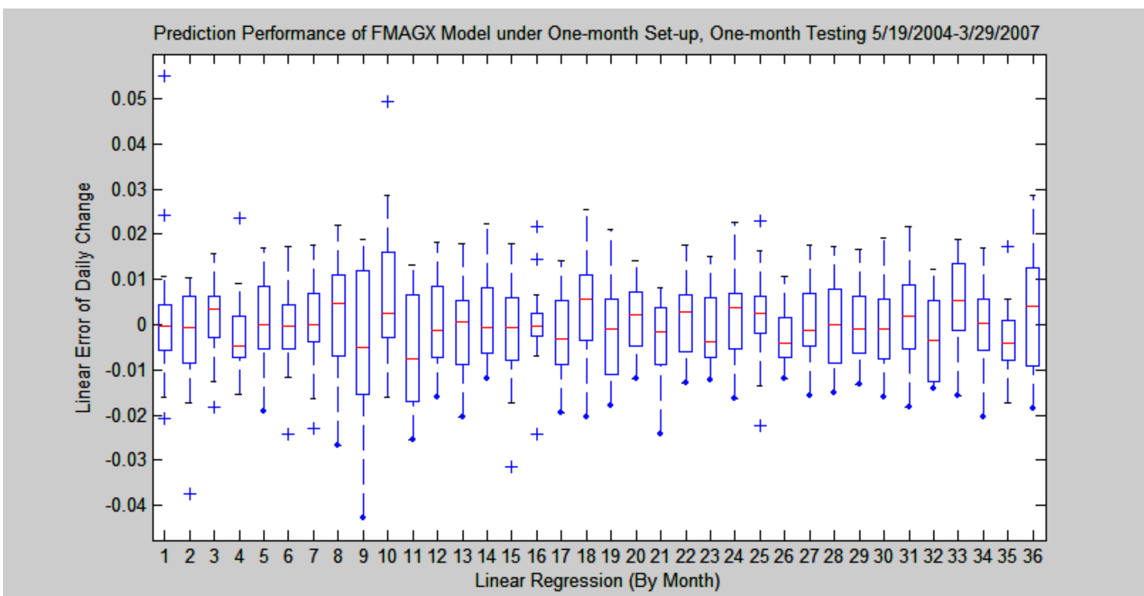
Exchange Traded Funds Description Table

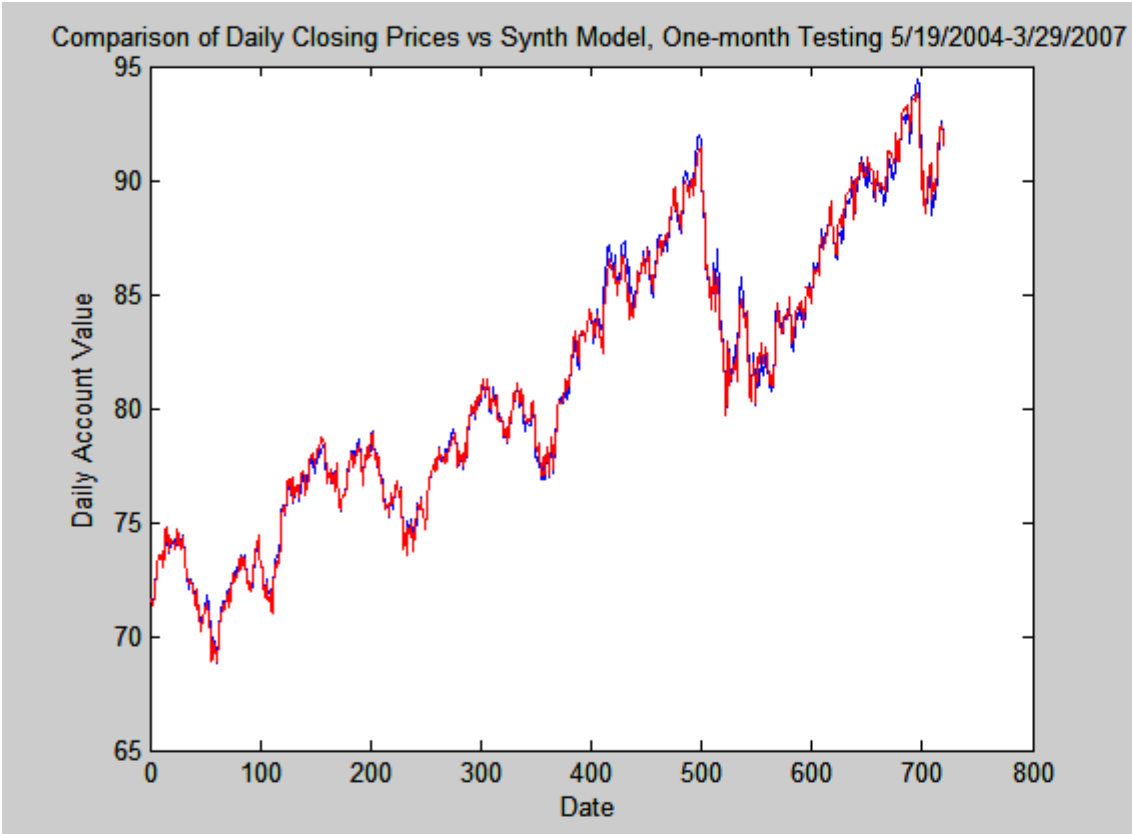
Breakdown of Various Indices Used [11]

SP500 *ETF: SPY*	The S&P 500 is an index containing the stocks of 500 Large-Cap corporations, most of which are American. The index is the most notable of the many indices owned and maintained by Standard & Poor's, a division of McGraw-Hill. S&P 500 is used in reference not only to the index but also to the 500 actual companies whose stocks are included in the index. The S&P 500 index forms part of the broader S&P 1500 and S&P Global 1200 stock market indices. All of the stocks in the index are those of large publicly held companies and trade on major US stock exchanges such as the New York Stock Exchange and Nasdaq. After the Dow Jones Industrial Average, the S&P 500 is the most widely watched index of large-cap US stocks. Float adjusted, market-value weighted.
Dow Jones Industrial Average *ETF: DIA*	the average consists of 30 of the largest and most widely held public companies in the United States. The "industrial" portion of the name is largely historical — many of the 30 modern components have little to do with heavy industry. To compensate for the effects of stock splits and other adjustments, it is currently a scaled average, not the actual average of the prices of its component stocks — the sum of the component prices is divided by a divisor, which changes over time, to generate the value of the index. The DJIA is criticized for being a price-weighted average, which gives relatively higher-priced stocks more influence over the average than their lower-priced counterparts. For example, a $1 increase in a lower-priced stock can be negated by a $1 decrease in a much higher-priced stock, even though the first stock experienced a larger percentage change. Additionally, the inclusion of only 30 stocks in the average has brought on additional criticism of the average, as the DJIA is widely used as an indicator of overall market performance.
NASDAQ-100 *ETF: QQQQ*	The NASDAQ-100 is a stock market index of 100 of the largest domestic and international non-financial companies listed on the NASDAQ stock exchange. It is a modified market value-weighted index; the companies weights in the index are based on their market capitalization, with certain rules capping the influence of the largest components. It does not contain financial companies, and includes companies incorporated in Canada (e.g. Research In Motion), Israel (e.g. Check Point), India (e.g. Infosys), Singapore (e.g. Flextronics), Sweden (e.g Ericsson), Switzerland (e.g. Logitech) and Ireland (e.g.Ryanair); both of these factors differentiate this index from the S&P 500.
Energy Index *ETF: XLE*	The Energy Select Sector SPDR Fund (the Fund) is an index fund that seeks to replicate the total return of the Energy Select Sector Index of the Standard & Poor's 500 Composite Stock Index (S&P 500 Index). During the fiscal year ended September 30, 2004 (fiscal 2004), the Fund had a return of 48.27%, as compared to the Energy Select Sector Index return of

	48.91% and the S&P 500 Index return of 13.87%. The Fund invests in industries, such as energy equipment and services, and oil and gas services, among others. In fiscal 2004, its top five holdings were Exxon Mobil Corp., ChevronTexaco Corp., ConocoPhillips Inc., Schlumberger Ltd. and Occidental Petroleum Corp.
Semiconductor Index *ETF: SMH*	The Semiconductor HOLDRS Trust issues depositary receipts called Semiconductor HOLDRS, representing an undivided beneficial ownership in the United States-traded common stock of companies that develop, manufacture and market integrated circuitry and other products known as semiconductors, which allow for speed and functionality in components used in computers and other electronic devices. The Bank of New York is the trustee. The 20 issuers of underlying securities represented by a Semiconductor HOLDR, as of August 1, 2005, were Analog Devices, Inc., Altera Corporation, Applied Materials Inc, Advanced Micro Devices, Inc., Amkor Technology Inc, Atmel Corp, Broadcom Corp, Intel Corporation, KLA-Tencor Corporation, Linear Technology Corporation, LSI Logic Corporation, Micron Technology Inc., Maxim Integrated Products Inc, National Semiconductor Corporation, Novellus Systems Inc, SanDisk Corporation, Teradyne, Inc., Texas Instruments Incorporated, Vitesse Semiconductor Corp and Xilinx Inc.

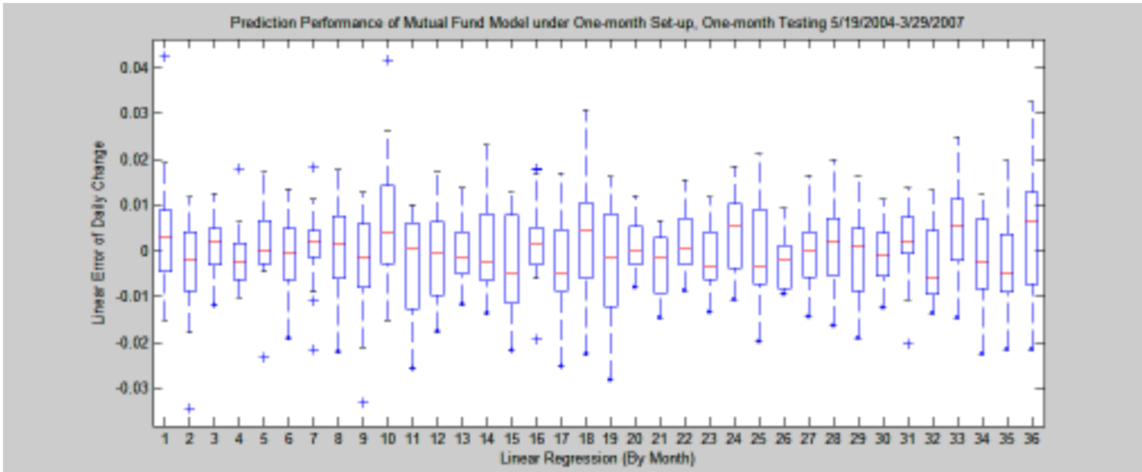
THREE-YEAR FMAGX RESULTS

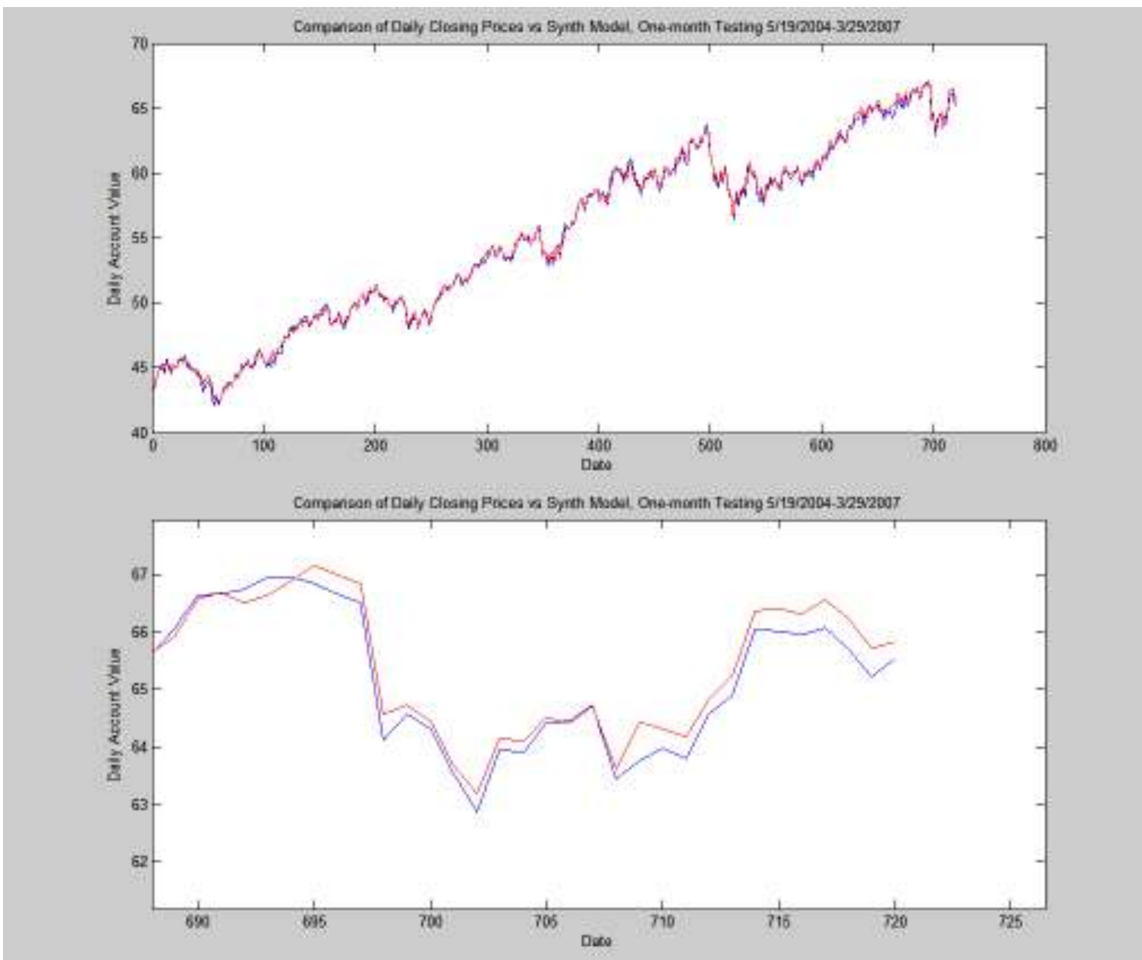


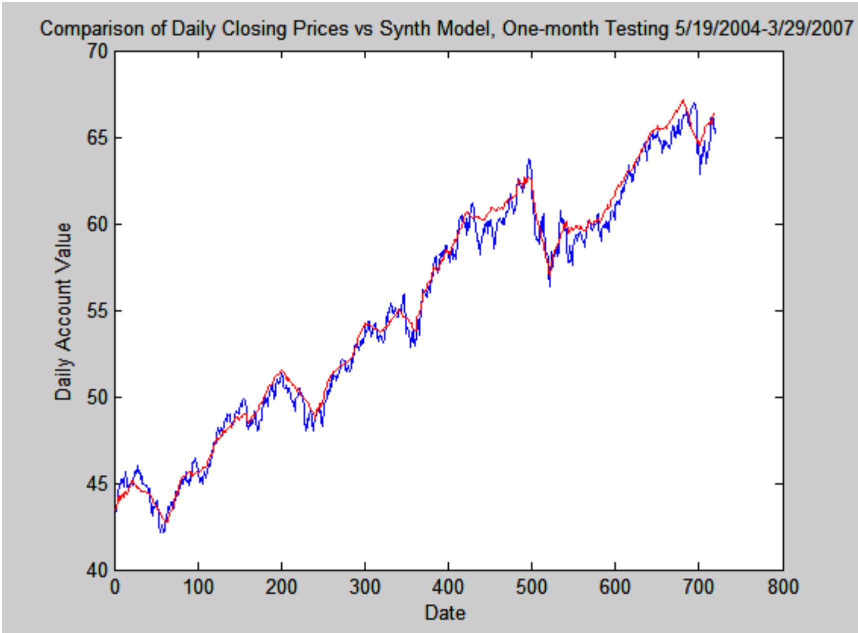


FMAGX, 1-Month Train Coefficients 20 April 04 through 29 March 2007			
Coefficient	**w0**	**SP500**	**DJIA**
Avg.	0.00	1.27	-0.04
Sdv.	0.00	0.57	0.47

THREE-YEAR FCNTX RESULTS



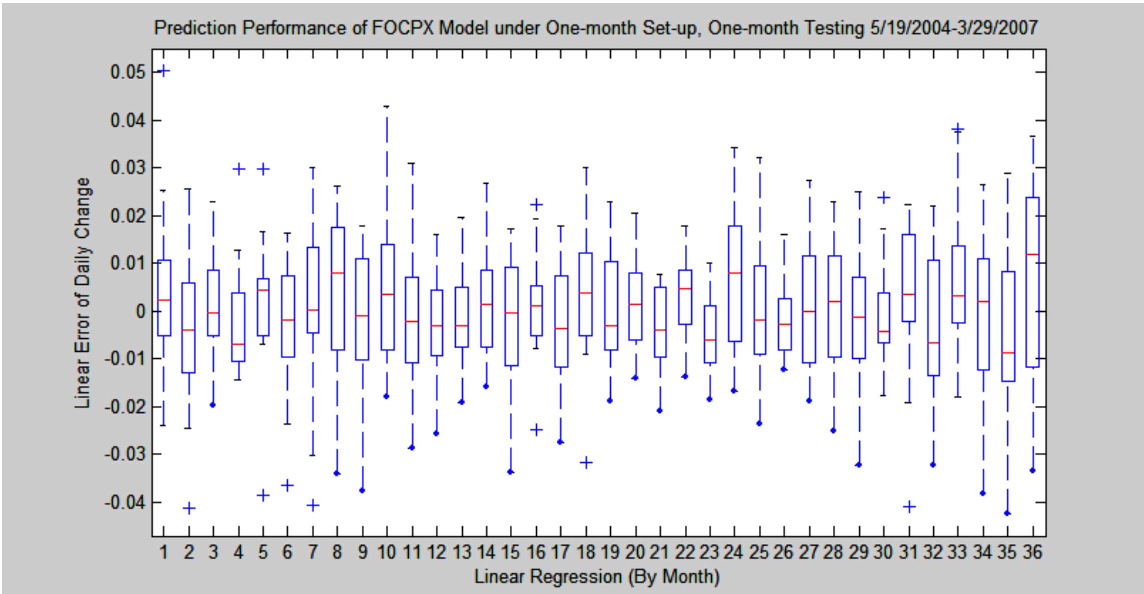


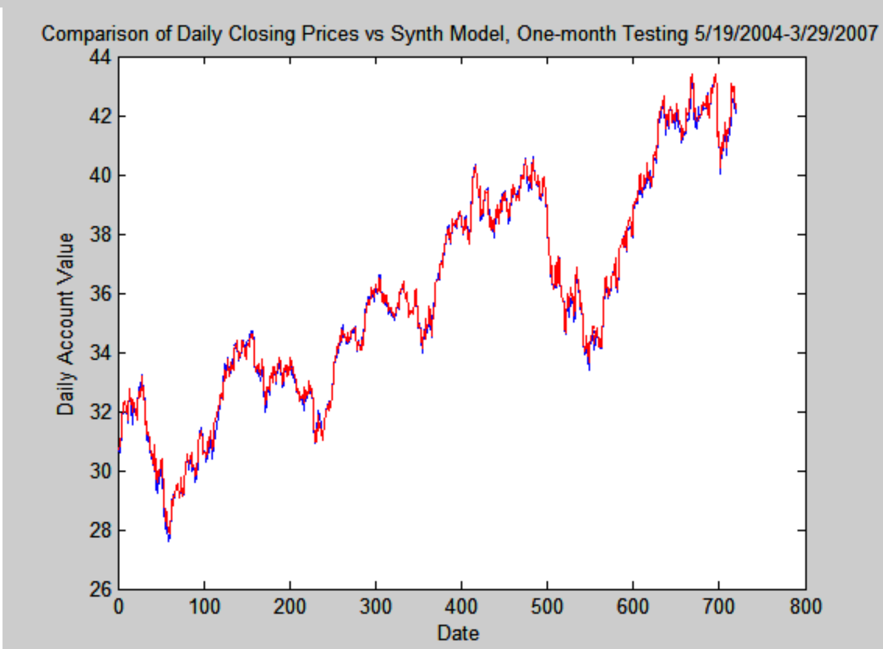


As shown above, synthetic-FCNTX with input variables SPY and DIA has a much smoother prediction model which does not as accurately reflect the true price changes of FCNTX as does a synthetic model with input variables XLE and QQQQ.

FCNTX, 1-Month Train Coefficients			
20 April 04 through 29 March 2007			
Coefficient	**w0**	**SP500**	**XLE**
Avg.	0.00	0.06	0.33
Sdv.	0.00	0.17	0.17

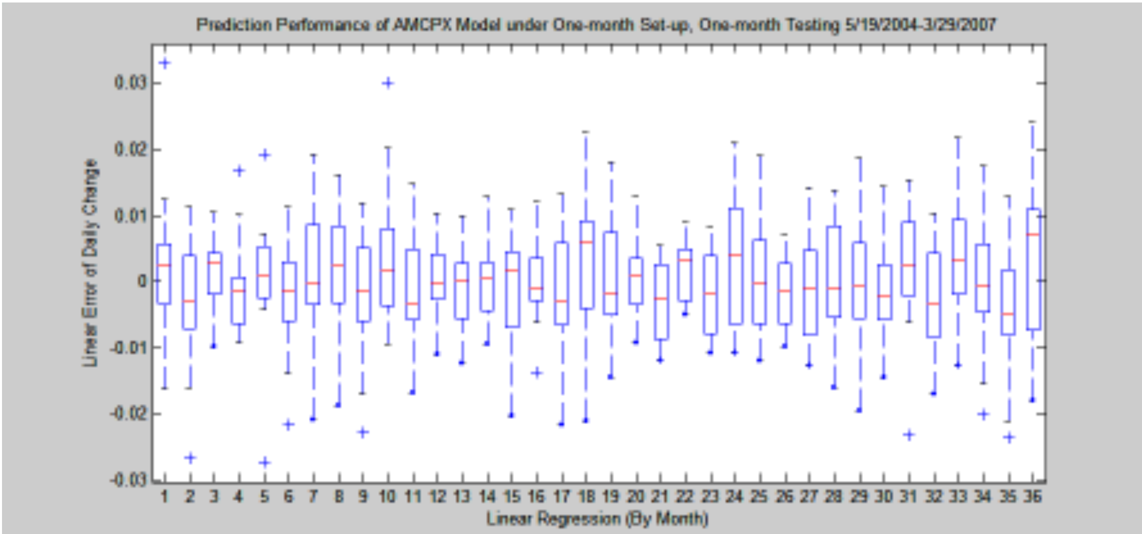
THREE-YEAR FOCPX RESULTS

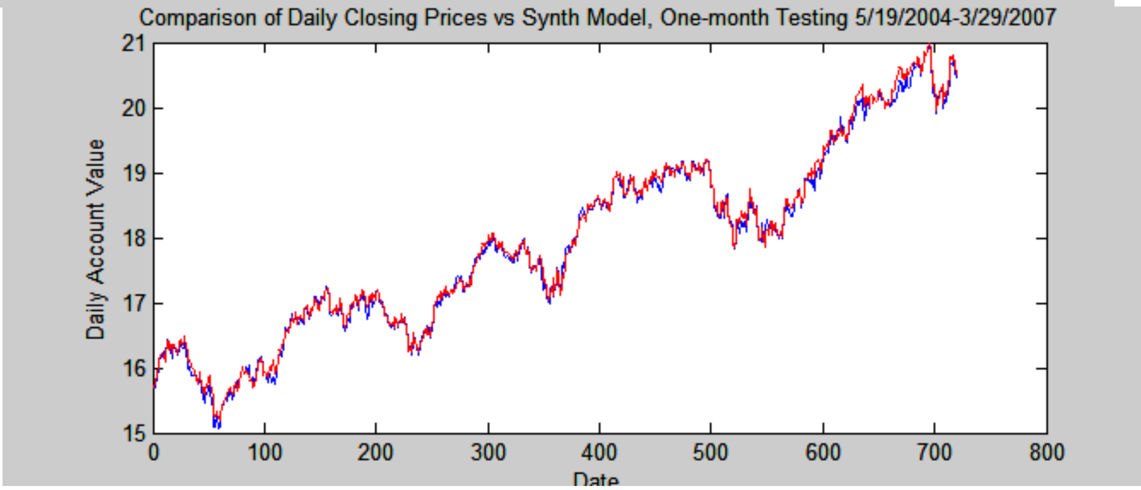




FOCPX, 1-Month Train Coefficients				
20 April 04 through 29 March 2007				
Coefficient	**w0**	**SP500**	**QQQQ**	**SMH**
Avg.	0.00	0.08	0.94	0.057
Sdv.	0.00	0.11	0.11	0.091

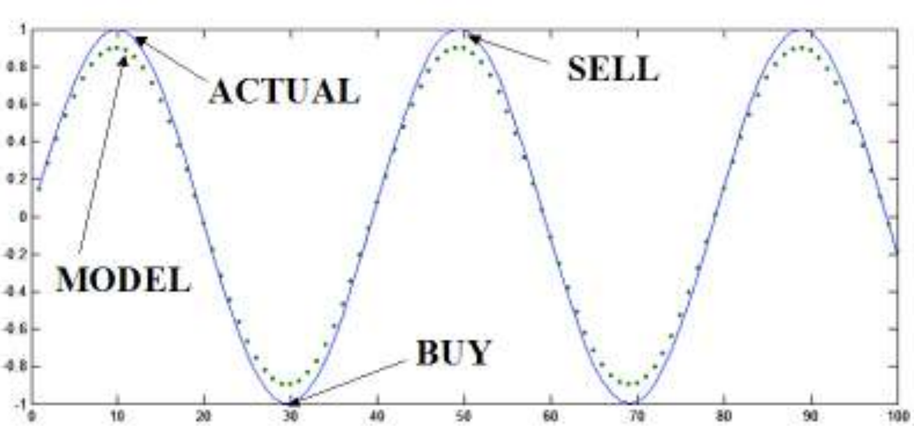
THREE-YEAR AMCPX RESULTS



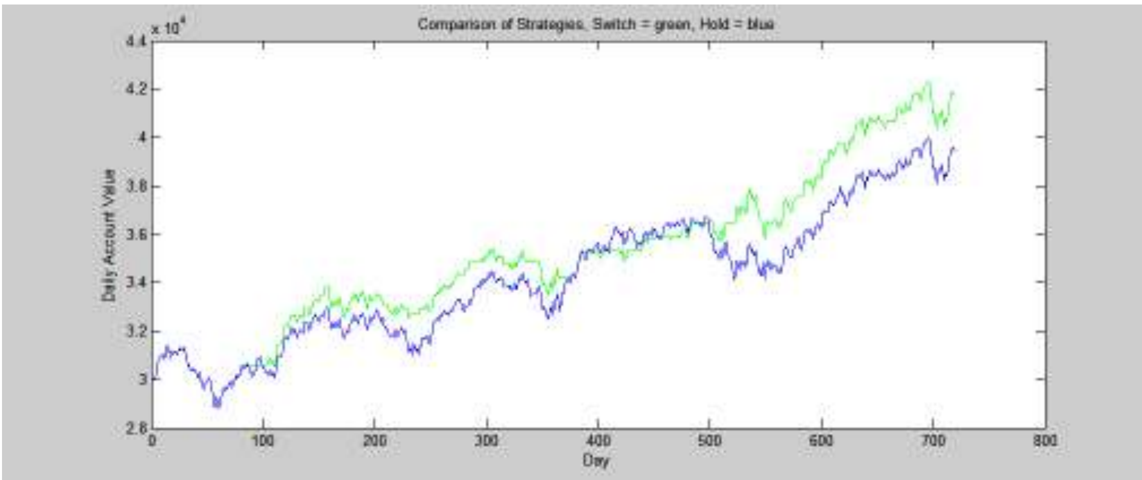


AMCPX, 1-Month Train Coefficients 20 April 04 through 29 March 2007			
Coefficient	**w0**	**SP500**	**QQQQ**
Avg.	0.00	0.007143	0.580517
Sdv.	0.00	0.071697	0.089823

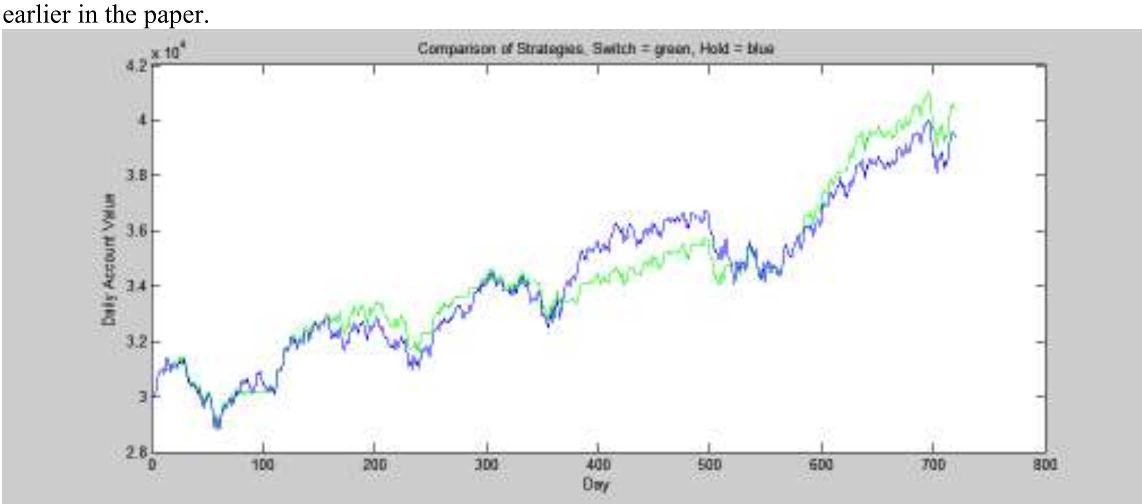
BUILDING A TRADING STRATEGY BASED ON AMCPX



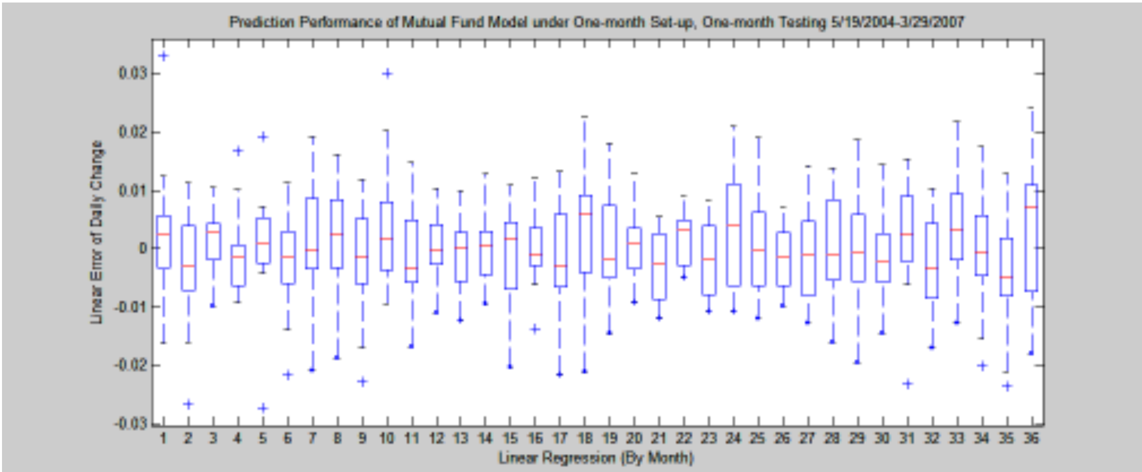
$$\frac{\sum_{i}^{-n}\left(NAV_i^{synthetic} - NAV_i^{actual}\right)}{n} = \delta_{n\_day\_average}$$



"Switch Strategy" (green) compared to "buy and hold" strategy (blue) with accurate AMCPX model found earlier in the paper.



"Switch Strategy" (green) compared to "buy and hold" strategy (blue) with accurate AMCPX model found with less accurate synthetic model.



Box plots for less accurate synthetic model for AMCPX (compared to more accurate model found earlier in the paper).

REFERENCES

[1] Perritt, Gerald W. *Mutual Funds Made Easy!* Chicago: Dearborn Financial Publishing Inc., 1995. 1.

[2] About ICI. Investment Company Institute (ICI). Retrieved on April 11, 2006.

[3] Zitzewitz, Eric. *Who Cares About Shareholders? Arbitrage-Proofing Mutual Funds.* The Journal of Law, Economics & Organization, V19 N2. Oxford University Press (2003)

[4] Hui Gao and Vladimir Cherkassky. Real-Time Pricing of Mutual Funds. 2006 International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada July 16-21, 2006.

[5] http://personal.fidelity.com/fidbin/getewdoc.cgi?fund=21

[6] Google Finance, Yahoo Finance Descriptions

[7] Google Finance, Yahoo Finance Descriptions

[8] Google Finance, Yahoo Finance Descriptions

[9] Fama, Eugene F. *The Behavior of Stock Market Prices*. Chicago: The Journal of Business Vol. 38, No. 1 (Jan., 1965), pp. 34-105

[10] Zitzewitz, Eric. *Who Cares About Shareholders? Arbitrage-Proofing Mutual Funds.* The Journal of Law, Economics & Organization, V19 N2. Oxford University Press (2003)

[11] Google Finance, Yahoo Finance Descriptions