WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia IEEE CEC

Using Sector Information with Linear Genetic Programming for Intraday Equity Price Trend Analysis

Garnett Wilson, *Member, IEEE*, Derek Leblanc, and Wolfgang Banzhaf

***Abstract*— A number of researchers who apply genetic programming (GP) to the analysis of financial data have had success in using predictability pretests to determine whether the time series under analysis by a GP contains patterns that are actually inherently predictable. However, most studies to date apply no such pretests, or pretests of any kind. Most previous work in this area has attempted to use filters to ensure inherent predictability of the data within a window of a time series, whereas other works have used multiple time frame windows under analysis by the GP to provide one overall GP recommendation. This work, for the first time, analyzes the use of external information about the price trend of a stock's market sector. This information is used in a filter to bolster confidence of a GP-based alert regarding formation of a trend for the chosen stock. Our results indicate a significant improvement in trend identification for the majority of stocks analyzed using intraday data.**

I. Introduction

Most researchers who apply evolutionary computation (EC) to the analysis of equity trends believe that there exist opportunities to identify, and take advantage of, patterns that indicate that the price of an equity or other financial instrument will rise or fall in the near future. The fact remains, though, that a handful of genetic programming (GP) practitioners remain concerned about the predictive ability of GP systems for financial analysis[2], [3]. These authors have worked to determine whether or not there is actually an underlying pattern in any price time series that will be analyzed by their GP system. The greatest benefit of these pre-emptive measures is that they potentially reduce or prevent unprofitable trading signals of a GP system in virtue of that system having performed search on an inherently unpredictable time series. In practice, this often means that the GP system becomes more reserved with its trading signals in an attempt to produce more profitable trading decisions. Most filters use inherent information about a window of the time series itself to determine whether or not it contains an inherently predictable pattern, or uses examination of sections of the time series or shuffled variants of it by a GP to determine predictability. In this paper, we introduce a filter that uses outside information to bolster the confidence of GP-based trend identification for an equity price series. In particular, we examine the effect of using time series price data from an exchange-traded fund (ETF) that represents the sector corresponding to the equity we wish to examine.

The remainder of this paper is organized as follows. Section 2 examines current systems that involve the measurement of predictability and use of filters, and Section 3 provides an overview of the LGP trading system using an ETF based on an appropriate market sector to create filters for GP trading decisions of an equity. Section 4 provides results that examine the effectiveness of the ETF-based filters with respect to trading behavior and overall profitability, and conclusions follow in Section 6.

Garnett Wilson, Derek Leblanc, and Wolfgang Banzhaf are with Afinin Labs Inc., St. John's, NL, Canada (email: gwilson@afinin.com, dleblanc@afinin.com, banzhaf@afinin.com).

II. Background

The first researcher to examine the effect of pretesting data used by a GP system for financial analysis was Kaboudan [2], where he introduced the η statistic. The η statistic measures predictability of a time series by comparing the results of a GP run on the unaltered time series of prices to results using a GP run on a shuffled version of the same time series. Kaboudan applied this technique to eight Dow Jones stocks and determined that predictability had an inverse relationship with length of period between price ticks. Chen et al. [3], instead of focusing on the inherent information about the time series itself, compare the results of GP to both random search and a search method called "lottery" trading that uses the outcome of a random variable to guide search. Chen et al. found that using these methods successfully showed increase in predictive ability for their GP system in a portion of the nine markets they tested. The authors of this work have previously used filters designed to be computationally efficient to examine inherent time series information for an LGP system in [5]. They found that a high frequency filter that was used consistently outperformed a basic filterless system on the chosen stock trends. The authors have also used a GP itself, applied to multiple time frames, to improve the trend identification abilities of GP in [6]. Using this technique, the authors determined that an increased number of time frames in a filter generally led to more conservative trading behavior, sometimes at the cost of missing profitable trading opportunities. Other related work includes that of Li and Tsang [4], where a specialized form of genetic programming with decision trees of rule sets called "FGP" was restricted by specifying the minimum and maximum of recommendations that the system could make on training data. The authors found that the failure rate of the GP could be reduced, but accuracy had to be maintained at the expense of an increased number of missed opportunities.

III. Linear GP Trading System Using Sector-based ETFs for Trend Identification

We apply a linear genetic programming (LGP) implementation on a minute-to-minute basis to intraday equity price

978-1-4673-1509-8/12/$31.00 ©2012 IEEE

trends using pretests, or "filters," corresponding to the price movement of a related ETF reflecting the movement of the market sector to which the equity belongs. The system is rule-based, with discrete trading alert-based outcomes. GP individuals are sets of trading rules combined with internal registers and a dedicated output (trade) register for storing a trade signal that results from execution of the instruction set. The LGP function set used to create the individual rules of the rule set includes standard mathematical operators (+, -, *), logical operators ($<$, $>$, =), and a number of established financial technical indicators including moving average, momentum, Bollinger bands, and current high, low, open, or close prices. There is an assumed per trade transaction cost of $10. Each LGP tournament is 1000 rounds long. With respect to GP operators, XOR mutation on a (uniform) randomly chosen instruction is used after each round at a rate of 0.5, and crossover is used each round with a rate of 0.9. Data used for a trading decision were normalized through two-phase preprocessing similar to the process described in [1]: Each price value was transformed by division by a lagged moving average, and then normalized with linear scaling into the range [0, 1] using

$$v_{scaled} = \frac{v_t - l_n}{h_n - l_n}$$

where v_{scaled} is the normalized trading value, v_t is the transformed trading value at time series data point t, h_n is the largest transformed value in the last n data points, l_n is the lowest transformed value in the last n data points, and n is the length of the lag selected for the initial transformation.

The best individual produced by the LGP is used in a "live" trading system. For live trading, information for prices m to n in the past is provided to the LGP by the live trading system. The LGP then returns a trading signal to the live trading system, which it uses to trade on the following time step, n + 1. LGP search was conducted to provide a buy, sell, or hold signal to the "live" trading system, which then places a trade of up to $1,000,000 based on its existing assets.

The first type of filter on top of the linear GP system that we consider is denoted "Sector Price Trend" and works simply by only allowing a buy from the underlying LGP for an equity if the line of best fit slope of the price data for the corresponding sector ETF in the past 5 minutes is positive. The second type of filter is denoted "Sector GP Alert" and only allows a buy from the underlying LGP on an equity if LGP has issued a buy for the corresponding sector ETF in the past 5 minutes. Thus, one filter is based solely on the price information of an external financial instrument, while the other filter is based on additional GP analysis (only the GP is analyzing a financial instrument other than the equity itself).

IV. Results

The LGP system examined intraday last sale stock price data from January 7, 2012 (chosen arbitrarily) for four stocks and their associated sectors from either the Nasdaq or NYSE exchanges: AAPL (Apple Inc.) in the technology sector, BBBY (Bed Bath & Beyond Inc.) in the services sector, MYL (Mylan Inc.) in the health care sector, and RY (Royal Bank of Canada) in the financial sector. The total of time series price data points for the day is 390 (6.5 trading hours at 60 prices per hour). The sector information at any given time is taken from the price data of an ETF from the SPDR®(Standard & Poor's depositary receipt) fund group from State Street Global Advisors that samples a significant number of equities from that sector: XLK for the technology sector, XLY for the services sector, XLV for the health care sector, and XLF for the finance sector. The stocks were selected to provide different trend types to test the filters, and they are not based on any fundamental information about the traded company. The first section of the results section discusses the performance of GP systems throughout the day-long price time series, and the second section discusses the profitability of the implementations after the close of markets. Multiple trials indicate little to no variation in trading decision across experiments for our GP trading system, so only one typical run need be shown.

A. *Performance for Price Series*

The performance of the buy-and-hold strategy is provided as a baseline for both the equity and the ETF in the top two graphs of each figure, where the maximum number of shares is purchased on the first day and held for the entire 390 minute trading day. Each trading scenario used an initial $1,000,000 with which to trade for the day, and the algorithm could trade up to that amount at any given time. The total value of each implementation's resources in cash and current cash value of total shares at each minute is plotted in the top two graphs of Figures 1, 3, 5, and 7. The performance of the LGP algorithm trading only the equity in question and its corresponding sector ETF are then analyzed in the two graphs just below the buy-and-hold graphs. The performance using only the equity shows how simple trading of that equity would occur without the influence of the ETF-based filters, and the performance of the ETF shows its influence for the sector-based filters. Finally, in the bottom two graphs of Figures 1, 3, 5, and 7 the equity traded by LGP using the sector based price trend filter and a sector-based GP alert filter are plotted. It is these graphs that test the novelty of the use of sector information in the trend prediction capabilities of the LGP system. We also examine the performance of the LGP system when used on the equity, the ETF without filters, and the equity traded by LGP under the influence of the two ETF filters as a ratio to buy-and-hold on the left of Figures 2, 4, 6, and 8. In addition, we also provide a plot to easily determine by a ratio whether or not LGP with each of the two types of ETF filters outperformed trading of the equity by LGP alone in the right graph of Figures 2, 4, 6, and 8. The overall worth results as described are shown below in Figures 1 to 8.

Examining Figure 1 (top two plots), we can see that when comparing the behavior of AAPL to the sector via the XLK ETF, we notice that the behavior of AAPL as an independent equity deviated considerably from the ETF. The trend we see

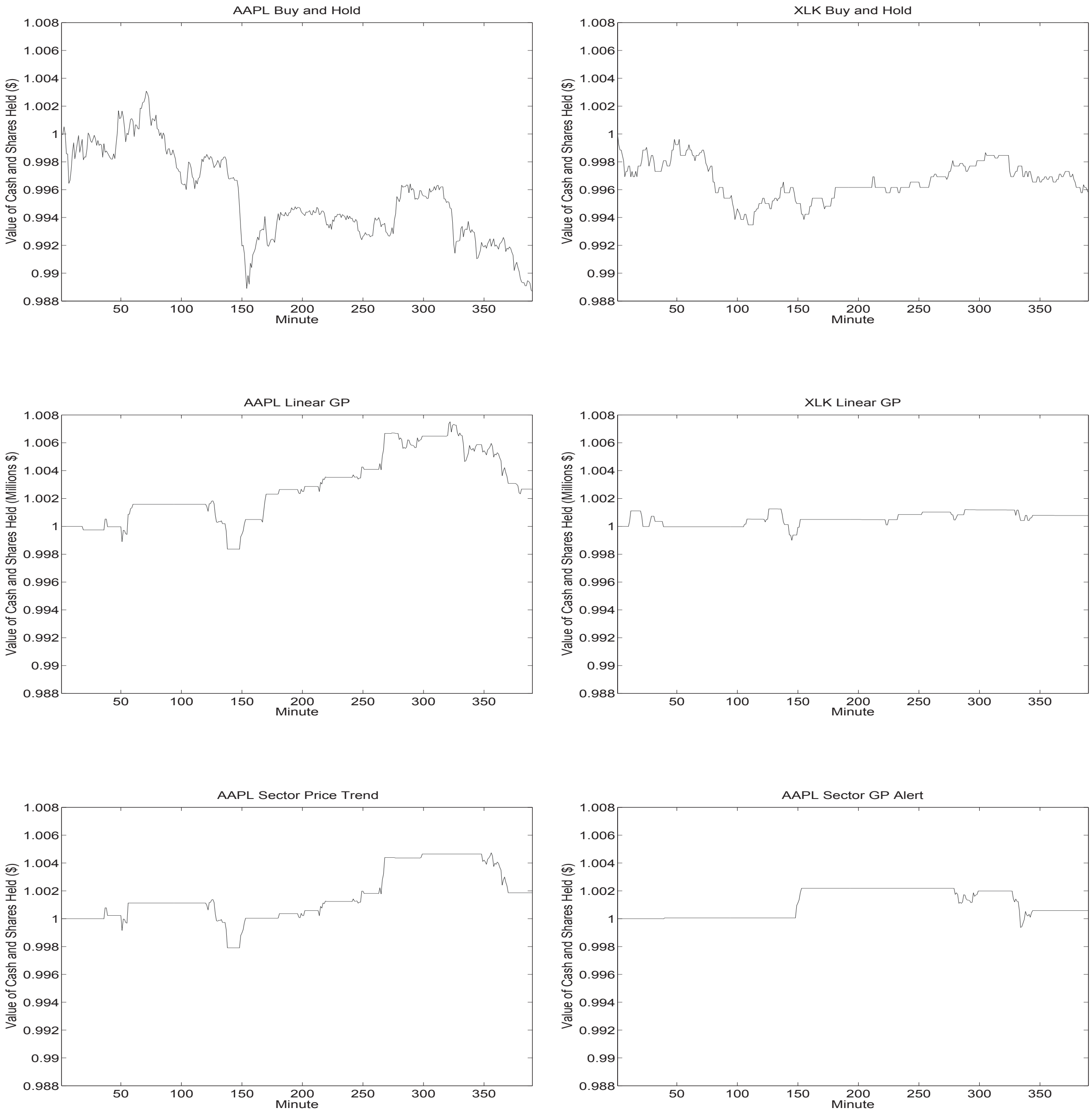

Fig. 1. Analysis of value of assets (cash and shares held) given initial $1,000,000 for AAPL and XLK as sector indicator.

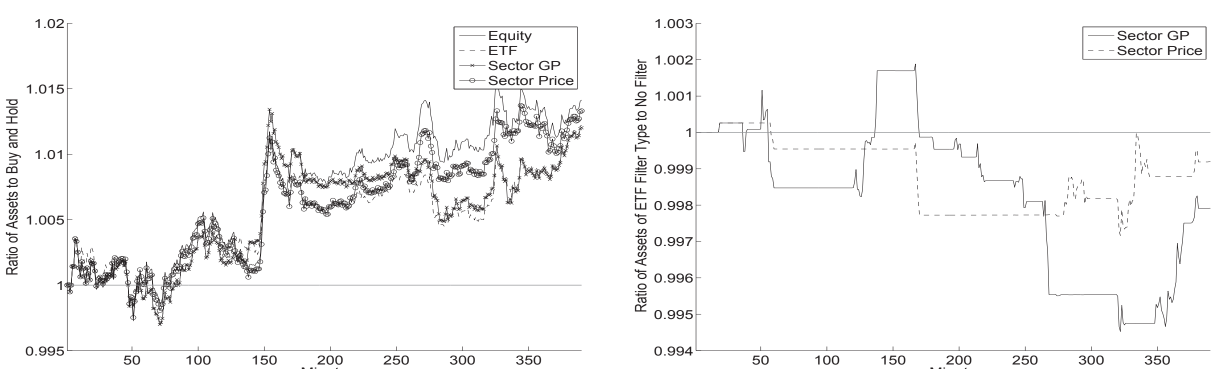

Fig. 2. Ratio-based analysis of value of LGP assets (cash and shares held) given initial $1,000,000 to buy-and-hold (left) and LGP only (right) for AAPL.

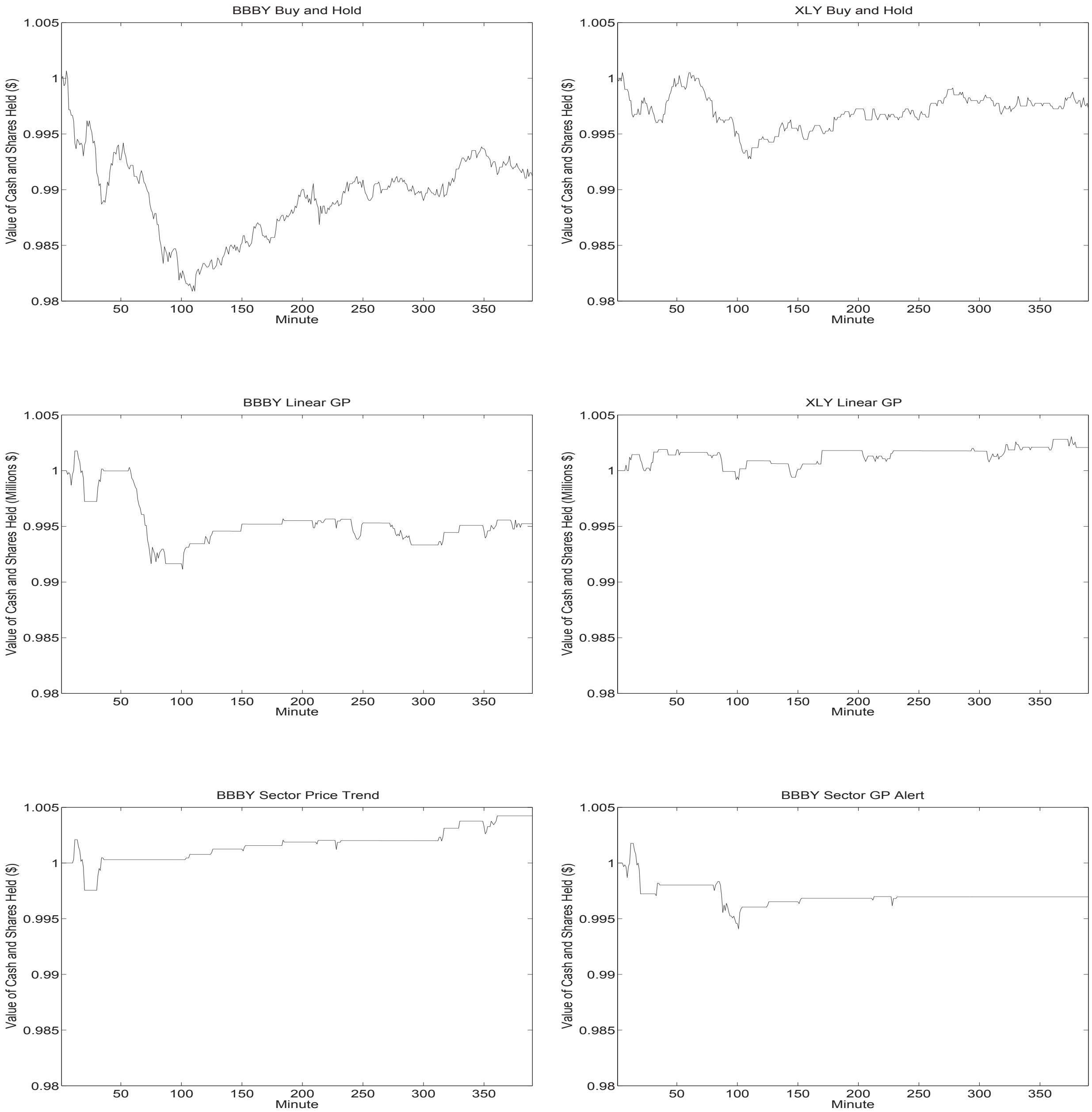

Fig. 3. Analysis of value of assets (cash and shares held) given initial $1,000,000 for BBBY and XLY as sector indicator.

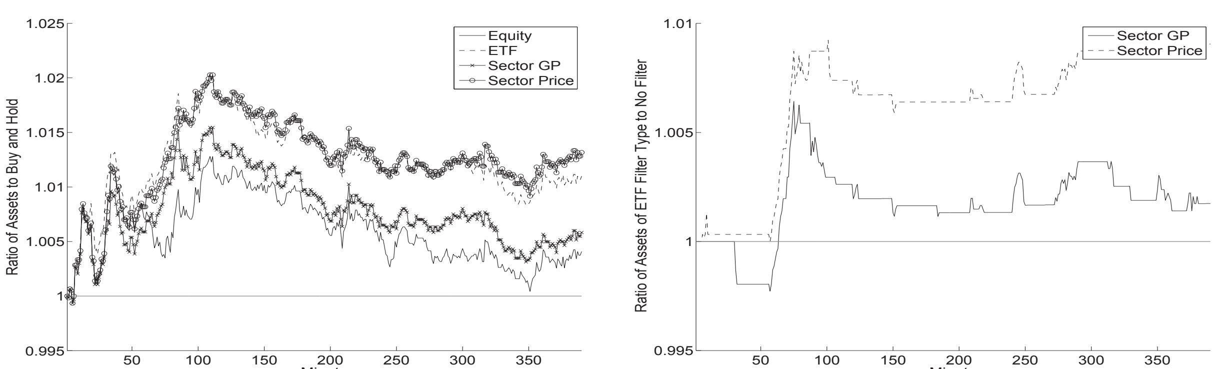

Fig. 4. Ratio-based analysis of value of LGP assets (cash and shares held) given initial $1,000,000 to buy-and-hold (left) and LGP only (right) for BBBY.

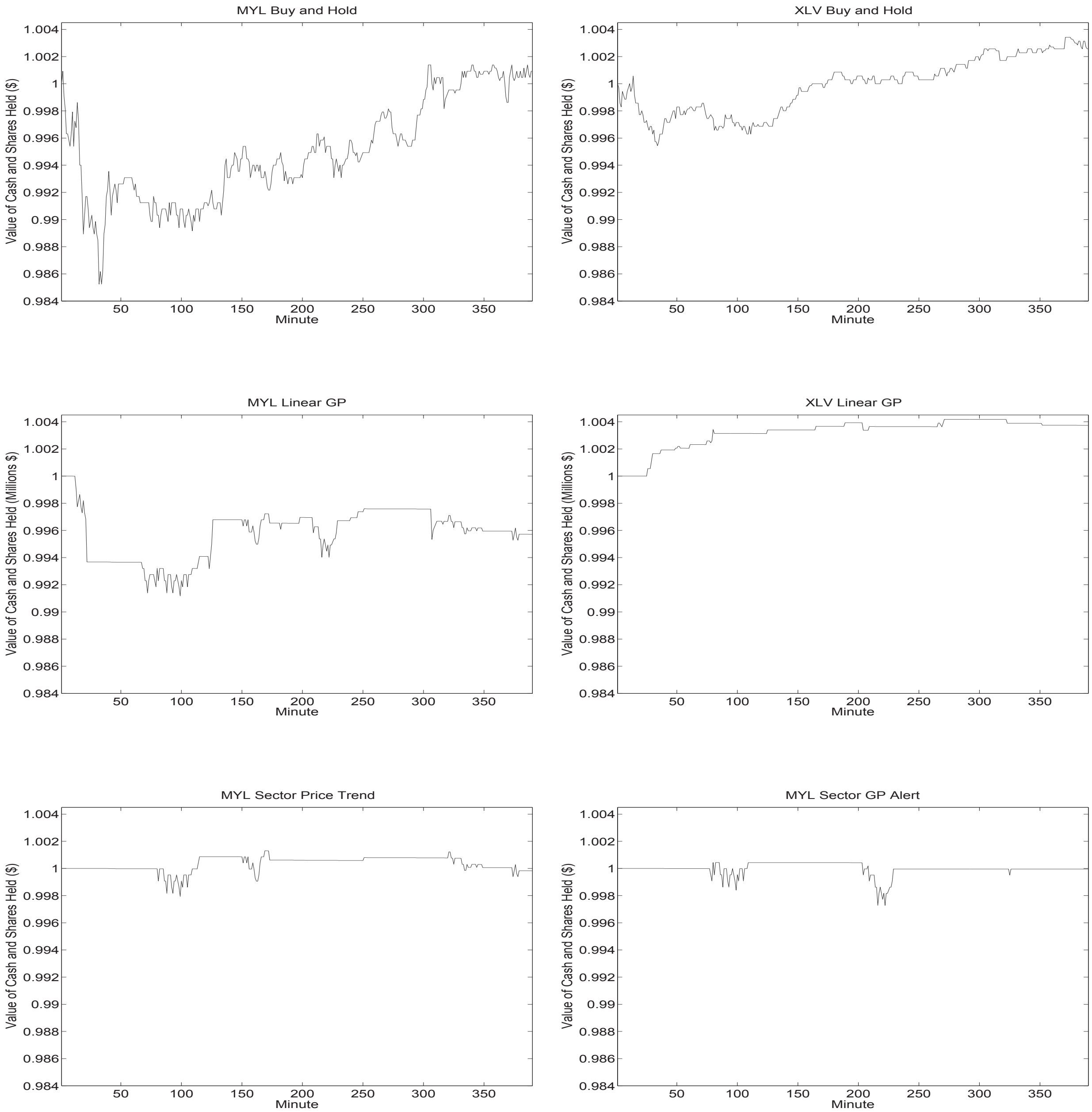

Fig. 5. Analysis of value of assets (cash and shares held) given initial $1,000,000 for MYL and XLV as sector indicator.

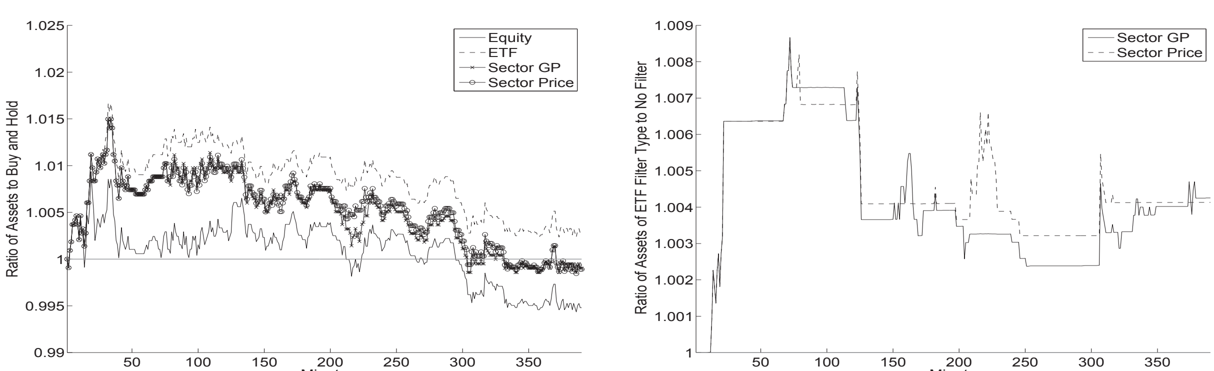

Fig. 6. Ratio-based analysis of value of LGP assets (cash and shares held) given initial $1,000,000 to buy-and-hold (left) and LGP only (right) for MYL.

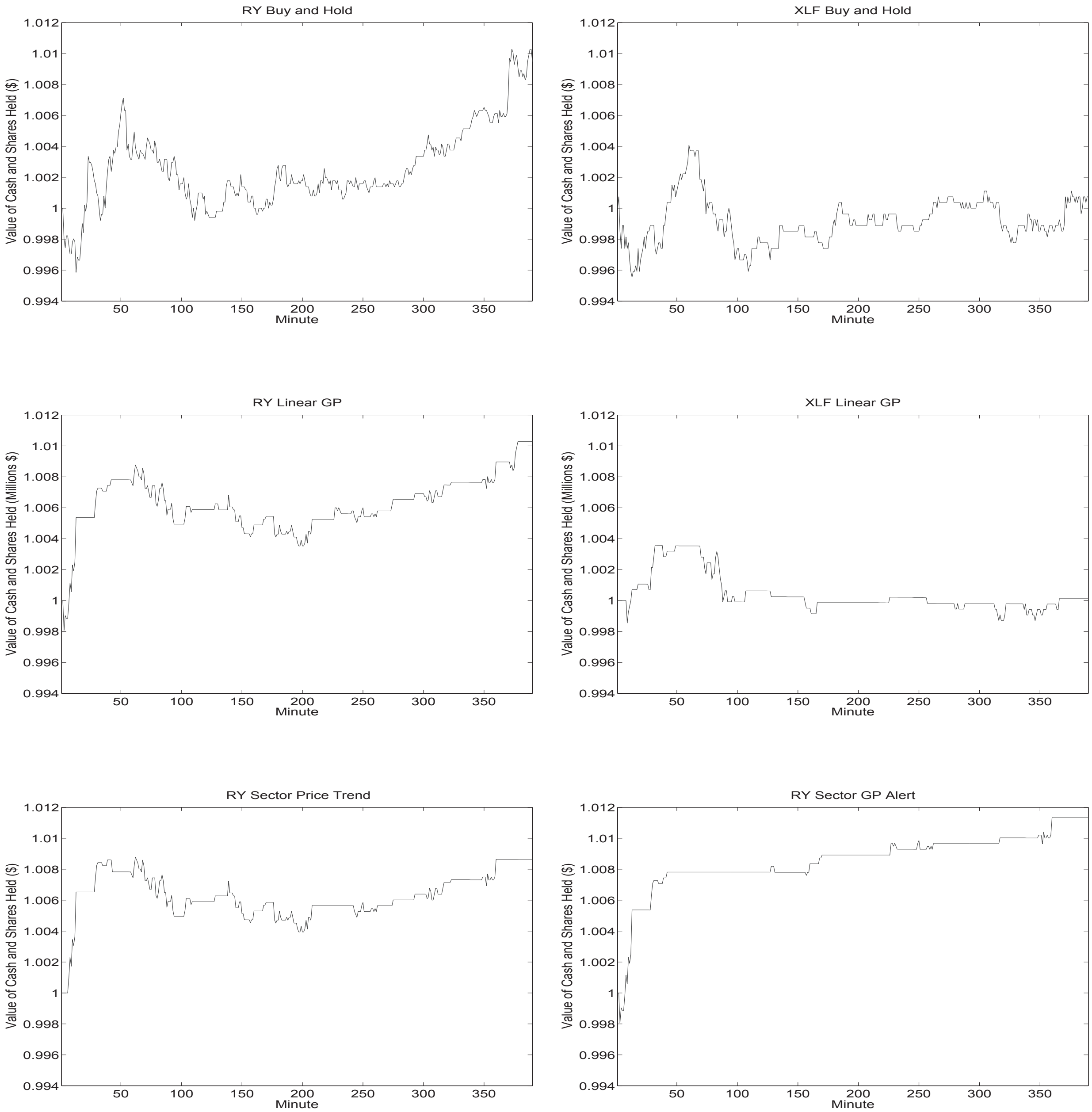

Fig. 7. Analysis of value of assets (cash and shares held) given initial $1,000,000 for RY and XLF as sector indicator.

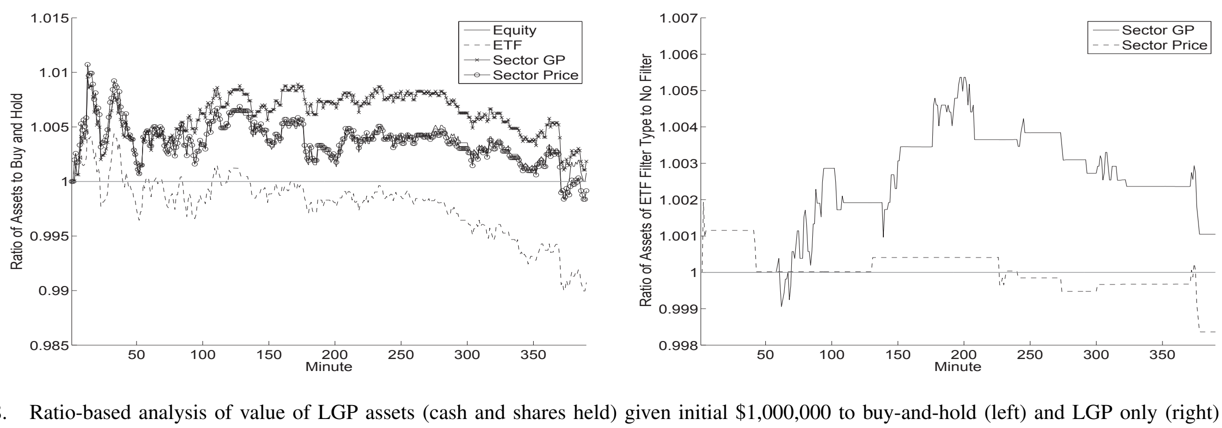

Fig. 8. Ratio-based analysis of value of LGP assets (cash and shares held) given initial $1,000,000 to buy-and-hold (left) and LGP only (right) for RY.

in AAPL as a lone equity is an overall downward price trend that is highly volatile, with a couple of instances of more outstanding downward spikes. Linear GP applied to AAPL generated much more profit than when it was applied to the sector ETF alone (middle two graphs). In the instance of AAPL, neither the use of the sector-based ETF as a filter using a slope, nor the action of LGP on the sector ETF outperformed the LGP alone for the AAPL equity (bottom two graphs). The comparative performance of the different implementations can be clearly compared using the ratio-based plots of Figure 2. It is evident from the leftmost plot, as just stated, that the ETF filters added to LGP do not provide an improvement when trading AAPL. This finding is clearly reinforced in the rightmost plot of Figure 2, which also shows that only during one period of time does the sector-based GP filter outperform simple use of the LGP on AAPL.

Figure 3 analyzes the equity BBBY, which exhibits an overall sideways trend after an initial downward trend for approximately the first 100 minutes of the day followed by a gradual rise for the rest of the day (upper left plot). The sector-based ETF used here, XLY, mainly exhibits a sideways trend throughout the day with a noticeable drop in price corresponding to BBBY from approximately minute 52 to minute 100 (upper right plot). We can see that using LGP results in a loss of assets at the end of this trading day (middle left graph), but the use of the slope-based price trend filter with LGP allows for a profit from BBBY at the end of the trading day (lower leftmost graph). However, the use of the ETF-based sector filter does not yield the same profitability at the end of this trading day (lower rightmost graph). Figure 4 (left plot) shows that a sector price based filter outperforms simple buy-and-hold of BBBY, as well as the use of LGP for trading BBBY. The rightmost plot of Figure 4 reaffirms that a sector price slope-based filter produces higher profits than either no filter or a sector-based GP filter throughout almost the entire time series (from approximately minute 60 onwards), and that both sector-based filters outperform the implementation with no filters throughout the major part of the trading day.

Figure 5 examines the trading of stock MYL (top leftmost graph). The behavior of MYL throughout the trading day is a largely upward trend following a sharp decline in the first 40 minutes after market open. The sector based ETF (XLV) shows a trend that mirrors that of MYL, but with a more conservative and sustained climb in price (top rightmost graph). Traded using only the linear GP algorithm, MYL does not close with the profit, whereas XLV has a 0.4% profit for the day (middle graphs). Neither the sector price trend filter nor the sector GP filter provide a significant end of day profit, and show minor fluctuations in profitability throughout the day (bottom two graphs). We can see in Figure 6 (left plot) that both the sector-based filters with LGP outperform the use of LGP alone on MYL. In the right plot of Figure 6 it is evident that, while not profitable at the end of this particular trading day, both implementations of the sector-based filter used with LGP outperform the use of LGP alone on MYL for almost the entire trading day.

Figure 7 (top leftmost plot) shows the buy-and-hold behavior for stock RY. This stock provides the opportunity for the analysis of a largely upward trend featuring steep price increases, volatile declines, sideways trading behavior, and finally a more gradual climb to end off the trading day. The corresponding sector based ETF, XLF, mirrors the behavior of RY for approximately the first 270 min. of the trading day (see top right plot). At that point, RY features a sustained climb in price that is not reflected by XLF. Using LGP for the trading of RY, we see that there are significant profits throughout the day, but these are not seen when trading XLF using LGP alone (middle plots). We can see that a more sustained profit-achieving trend is possible when using the sector GP-based filter (bottom right plot). The use of the sector price slope as a filter also provides a profitable trend throughout the day, but it is more volatile and less sustained (bottom left plot). Figure 8, left plot shows that the ETF under performs the equity in this instance. This is the first time that this relationship is evident–in all other cases the ETF largely outperformed the chosen equity (see Figures 2, 4, and 6). It is also evident from the right plot of Figure 8 that simple LGP, and the addition of the two sector-based filters, all outperformed buy-and-hold for RY. The right plot of Figure 8 clearly indicates that the sector GP filter performed better overall than the sector price filter.

B. *Profitability and Trading Activity*

The previous section examined the value of the assets held by the system at various points throughout the trading day. This analysis provided a way of examining the behavior of the system at any given minute, and at the end of the trading day. In this section we examine the overall behavior of the system by discussing profitability. A better way than final profit (often found in studies such as this) to measure profitability of the system is to look at cumulative profit, done by keeping an ongoing total of each minute's level of profit in terms of current assets held. Thus, we provide the cumulative profitability of each LGP implementation relative to buy-and-hold over all points in the trading day for each stock and system/filter combination in Figures 9 to 12. Bottom, middle, and top of boxes indicate lower quartile, median, and upper quartile values, respectively. If notches of boxes do not overlap, medians of the two sets of data differ at the 0.95 confidence interval. The symbol '+' denotes points from 1.5 to 3 times the interquartile range, and 'o' denotes points outside 3 times the interquartile range.

Figure 9 for AAPL shows that trading of the equity using LGP generates higher profit, with statistical significance, than trading of the corresponding sector ETF with LGP. However, in this case, neither filter implementation outperforms the LGP trading on the equity only. Recall that for this equity, it was outperformed by its sector ETF in buy-and-hold (Figure 1), so this likely contributed to these results in the case of AAPL. In all of the remaining boxplots for BBBY, MYL, and RY (Figures 10 to Figure 12), the two sector based filter applications outperform the equity alone. These results

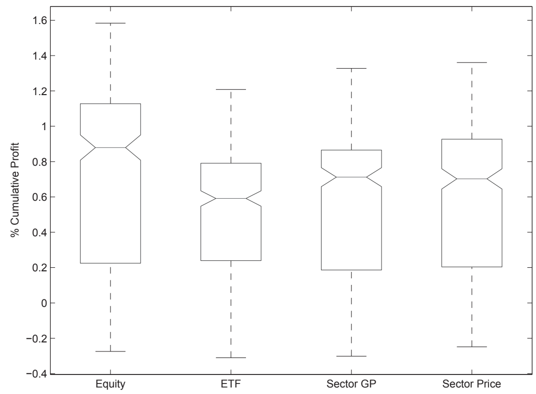

Fig. 9. Cumulative profit of LGP for AAPL, XLK, AAPL using XLK price filter, and AAPL using XLK GP filter.

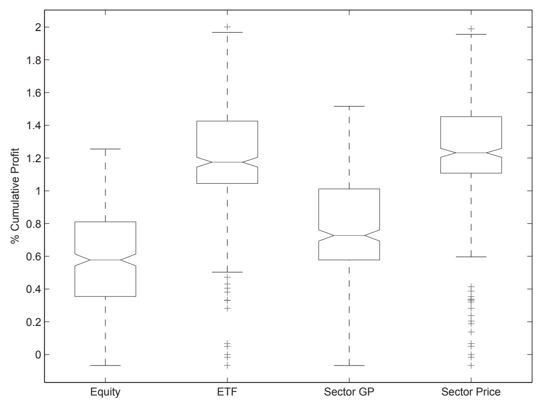

Fig. 10. Cumulative profit of LGP for BBBY, XLY, BBBY using XLY price filter, and BBBY using XLY GP filter.

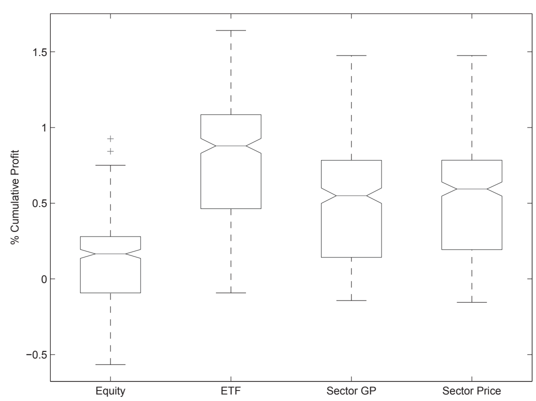

Fig. 11. Cumulative profit of LGP for MYL, XLV, MYL using XLV price filter, and MYL using XLV GP filter.

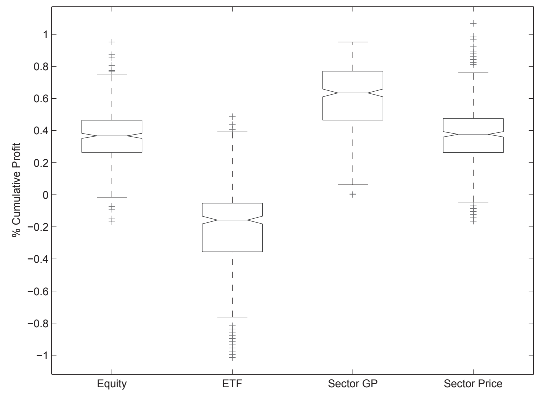

Fig. 12. Cumulative profit of LGP for RY, XLF, RY using XLF price filter, and RY using XLF GP filter.

are of statistical significance, except for Sector Price Filter compared to LGP for RY, where the general spread of the data indicates better performance using the filter. In the case of BBBY and MYL, the ETF outperformed the equity in buy-and-hold (see Figures 3 and 5). In the case of BBBY, the equity has grossly underperformed compared to its sector ETF in buy-and-hold (see Figure 3), so the LGP alone was naturally outperformed using the sector information. However, in all cases except for AAPL, we can see that any type of sector information provided as an additional filter to the basic LGP analysis of the equity is beneficial. To decide further which type of sector information (price data or GP analysis of that price data) is best for the filter in terms of cumulative profit, however, the data does not provide a clear indicator.

V. Conclusions

This work presents two predictability filters for an LGP system to provide market sector information through analysis of an ETF corresponding to the sector of the chosen equity. One type of sector filter examined raw price information about the ETF, while the other used LGP analysis on the ETF and provided that information to the underlying LGP operating on the equity. With the exception of one equity, there was benefit to analyzing the movement of an ETF for the sector corresponding to an equity. However, whether it is best to use the raw price trend of that ETF for the filter data or have the filter use GP on the ETF is unclear; sets of circumstances where one type of filter might make more sense than the other is a promising topic for future work.

References

[1] A. Brabazon and M. O'Neill, *Biologically Inspired Algorithms for Financial Modeling*, Berlin: Springer, 2006.

[2] M. A. Kaboudan, "A Measure of Time Series' Predictability Using Genetic Programming Applied to Stock Returns," *Journal of Forecasting*, vol. 18, pp. 345–357, 1999.

[3] S. Chen and N. Navet, "Failure of Genetic-Programming Induced Trading Strategies: Distinguishing between Efficient Markets and Inefficient Algorithms," in *Computational Intelligence in Economics and Finance*, S. Chen, P. P. Wang, and T. Kuo, Eds. Berlin: Springer, 2007.

[4] J. Li and E. Tsang, "Reducing Failures in Investment Recommendations using Genetic Programming," in *Sixth International Conference on Computing in Economics and Finance*, Society for Computational Economics, 2000.

[5] G. Wilson and W. Banzhaf, "Fast and Effective Predictability Filters for Stock Price Series using Linear Genetic Programming," in *2010 IEEE Congress on Evolutionary Computation (CEC)*, IEEE Press, 2010.

[6] G. Wilson, D. Leblanc, and W. Banzhaf, "Stock Trading using Linear Genetic Programming with Multiple Time Frames," in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, ACM Press, 2011.

[7] G. Wilson and W. Banzhaf, "Algorithmic Trading with Developmental and Linear Genetic Programming," in *Genetic Programming Theory and Practice VII*, R. Riolo, U. O'Reilly, and T. McConaghy, Eds. New York: Springer, 2009, ch. 8, pp. 119-134.