A MACHINE LEARNING APPROACH TO FORECASTING CONSUMER FOOD PRICES

by

Jabez (Jay) Harris

Submitted in partial fulfilment of the requirements

for the degree of Master of Electronic Commerce

at

Dalhousie University

Halifax, Nova Scotia

August 2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Building on the success of the Canada Food Price Report 2017 and its inclusion of a machine learning methodology, this research thesis posed and attempted to answer the following question, "What is the best way to predict food prices for the average Canadian consumer?" The Canada Consumer Price Index (CPI) was selected as the dependent variable and forecasted against three data models to access their predictive values. The models included the popular Holt-Winters Triple Smoothing Exponent model as a benchmark, a financial futures-market data model and a model adapted from the Canada Food Price Report 2017. The hope was to create a more robust forecast model for future Canada Food Price Reports and similar econometric predictions.

As hypothesized, the Financial Futures-Market based model outperformed the Food Price Report model with 1.6% and 2.4% average error rates respectively. Each model captured 4 of the 8 CPI food categories and apart from the CPI Seafood category in the case of the Food Price Report model, both models easily bested the popular Holt-Winters benchmark model.

The CPI Restaurant category, which was regarded as the most difficult to forecast because of its composite nature, produced the lowest error rate for both the Food Price Report and Financial Futures-Market models. Given the superb performance by both the Food Price Report and Financial Futures-Market models it is likely that even better results can be achieved by combining the datasets to share their uncommon information.

LIST OF ABBREVIATIONS AND SYMBOLS USED

ARIMA: Autoregressive Integrated Moving Average

CFS: Correlation-based Feature Selection

CPI: Consumer Price Index

ERS: Economic Research Service

FAO: Food and Agriculture Organization

MAPE: Mean Absolute Percentage Error

MSE: Mean Square Error

NAV: Net Asset Value

RMSE: Root Mean Square Error

US: United States of America

USDA: United States Department of Agriculture

WEKA: Waikato Environment of Knowledge Analysis

$\hat{y}_T$: Holt- Winters Forecasting equation

a_t: Holt-Winter Level equation

b_t: Holt-Winters Trend equation

s_t: Holt-Winter Season equation

α, β and, γ: Holt-Winters Smoothing Parameters

p: Holt-Winters Period Parameter

Y : Holt Winter Initial Value

ACKNOWLEDGEMENTS

To my supervisors, Dr. Vlado Keselj and Dr. Sylvain Charlebois: I am extremely grateful for the opportunity to work with you both on this research thesis. Your encouragement, advice and assistance throughout this experience has been a major driving force for my work. I appreciated your approach to supervising this research thesis and I truly hope you are proud of what I have accomplished.

Many thanks to my committee members, Dr. Carolyn Watters and Dr. Vladimir Lucic. Thank you for dedicating your time and efforts to ensuring that this research thesis aligns with the standards of academic research. Your contributions have helped me to realize a significant life goal and I certainly could not have achieved this without your help.

I am grateful to the staff of the Computer Science Department for their unfailing support and kindness throughout the course of my graduate program. I would also like to express my profound gratitude to NSERC, the Faculty of Management and the Faculty of Computer Science for providing access to funding opportunities for my graduate education.

Finally, to everyone at Shiftkey Labs, Dalhousie Academic Technology Services and the Centre for Learning and Teaching, the success that I have had while in Canada would have been impossible if not for your involvement. You gave me the opportunity to not only earn income but also to work with an amazing team of individuals who cared about me, respected my work and were genuinely interested in my success. I owe you my deepest gratitude.

CHAPTER 1: INTRODUCTION

1.1 Why forecast food prices?

1.1.1 The Rising Cost of Shelter, Transportation and Food

As of January 2015, the average Canadian household spent 62.3% of its total income on shelter, transportation and food costs alone (Statistics Canada, 2015). Though this estimate may seem outrageous, in our hearts Canadians know that it is indeed true. When we consider our current socio-economic situation, we realize that Canadians from all provinces and all walks of life have been affected by rising food costs, runaway housing prices and increasing transportation expenses.

Food Banks Canada's HungerCount 2016 report revealed that almost 75% of recipients of foodbank aid in Canada in 2016 had secured market-rate housing through rent or an established path to home ownership; however, to accomplish this many Canadians were unfortunately forced to trade between adequate housing and food security. This phenomenon, which has become prevalent in many developed countries, has led to a 28% increase in food bank visits in Canada since 2008, with 13% of Canadians living in a constant state of food insecurity (Food Banks Canada, 2016).

The National Bank of Canada's House Price Index estimated an increase of almost 13% in home prices in Canada's six largest metropolitan areas during the 12-month period from November 2015 to November 2016 (Teranet, 2016). In Vancouver, a single-family home priced below $1 million is now a rarity, as home prices continue to skyrocket amid an influx of foreign investors who view Canada as a gateway to previously inaccessible western markets and affluent lifestyles.

Accordingly, Statistics Canada (2016b) also reported in 2015, public transportation costs had soared to 36.5% higher than they were in 2002. To put this into context, adult bus fare in Toronto in 2002 could be purchased for $2 cash but in 2016 that same bus fare is now $3.25 cash for an increase of 62% in bus fares alone. This equates to an additional $50 in monthly expenses or $600 annually, since 2002, for a single person commuting to work daily. Similarly, for privately used transportation methods, gasoline prices in Toronto have risen by 50% since 2002 (Statistics Canada, 2016a). This represents an increasing burden on the Ontario consumer who still struggles to stay ahead of inflationary prices despite having experienced some wage increase since 2002 with a growth rate in average weekly wages of 40% (Statistics Canada, 2016c).

As mentioned above, in Toronto, Canada's most populous city, public and private transportation costs have outpaced the average weekly wage income. At this point it is not difficult to see how with shelter and transportation costs having soared to record breaking heights, why many Canadians have sought salvation in the substitutive characteristics of our favorite foods and the flexibility of retail food prices. In Canada, food prices represent a cost variable which can still be manipulated by consumers to circumvent rising costs in shelter and transportation.

The cauliflower "crisis" of 2016 illustrates how the Canadian marketplace is still very susceptible to the demands of its consumers. In January 2016, a New York Times article titled, "In Canada, the 8-Dollar cauliflower shows the pain of falling oil prices", highlighted how low commodity prices coupled with a devalued Canadian dollar and deficiencies in the supply chain had culminated in exorbitant food prices for Canadian shoppers. However, unlike similar events in the United States, Canadian shoppers - through social media-styled protests - managed to convince the major food retailers to absorb the excess costs being perpetuated by an unusually dry season in the farm intensive areas of California (Charlebois, et al., 2016).

Food prices are still a moving target and with careful planning and consideration it should be possible to endure the price hikes associated with other consumer goods and services by adjusting our food consumption habits. However, to confidently take advantage of the volatile food prices would require not only planning but also the ability to precisely forecast food price inflation and initiate appropriate stop-gap measures, such as increasing social aid, adjusting agricultural trade policies, pursuing new vendors, introducing price ceilings or addressing inefficiencies in distribution. We must also consider the global marketplace and Canada's vulnerability to the impact of rising global food prices given that since 2015 the country's food imports as a percentage of its merchandise imports are at the highest they have been since the 1970s (The World Bank, 2015). At the same time, Canada's food exports, as a percentage of its merchandise imports, are only now recovering from a 30-year low and near stagnant growth from 2010 to 2014 (Statistics Canada, 2016).

1.1.2 The Role of Inflation

Many economists see inflation as an unavoidable cost for maintaining a thriving and efficient economy but too low of an inflation rate could spell disaster for local businesses and economic policy. The proverbial sweet spot for inflation rates is believed to be between

1% and 2% annually and not 0% or even the occasional decline as many consumers would advocate for. The reasons for a positive and controlled inflation rate are (Billi & Kahn, 2008):

1) The available measures of inflation are not perfect and can overstate the "true" inflation rate.
2) A small amount of inflation may make it easier for firms to reduce real wages when necessary for maintaining employment levels.
3) A negative inflation rate due to lost incomes, or deflation, could be even more costly than a similar rate of inflation, which suggests that a low rate of inflation might be desirable to protect against the residual effects of falling prices.
4) At very low levels of inflation, nominal interest rates may be close to zero, limiting a central bank's ability to ease monetary policy in response to economic weakness.

All in all, a small amount of inflation is believed to be necessary to allow manufacturers and service providers to improve the quality of their wares annually without putting any undue burden on the consumer or the economy.

So, what exactly is inflation? Per Oner (2012), it is the measure of how much more expensive a set of goods and services have become over a certain period, usually one year. It is usually calculated with a broad scope such as the overall increases in consumer prices or cost of living for a specific sovereign nation (Öner, 2012). Alternatively, its scope can also be narrowed to focus in on individual price changes like food, services, education or fuel. In Canada, the Consumer Price Index is one of the majorly used measures of inflation and deflation (Statistics Canada, 1996). The CPI can be thought of as a measure of the percentage change over time in the average cost of a large basket of goods and services purchased by an average Canadian consumer. The basket contains commodities of goods and services that are kept constant over time in quantity and quality to ensure consistency across all applicable years and generations of Canadians. Thus, changes to the cost of the basket over time are not due to changes in the quantity or quality of the goods and services observed. Consequently, the index reflects only pure price movements (Statistics Canada, 1996).

However, inflation can distort and even reduce the purchasing power of the consumer over time (Öner, 2012). Consider, if inflation in a year was at 5%, pensioners who receive a fixed 5% yearly increase to their pension would be wary of their purchasing power as any further increase in the inflation rate over the annual pension increase rate would result in a reduction of their purchasing power and real income. On the other hand, a car owner who pays a fixed-rate automotive loan of 5% per annum could possibly benefit from 5% inflation, because the real interest rate, calculated as the nominal rate minus the inflation rate, would be zero. If the car owner's income can keep pace with the inflation rate, then servicing the debt will remain relatively easy.

The Car owner's example is even more intriguing because it hints to the price setting dilemma of having to forecast the rate of inflation to ensure that consumer prices remain at a profitable level. Of course, the price setter will suffer if the inflation rate was equal to or higher than the chosen price as this would result in a reduced income from the sales of that item. To offset any unforeseen changes to the economy the price setter will raise the price commensurate with the inflation cost. In this situation, the consumer suffers because of his inability to afford the item at the new price.

1.1.3 Forecasting Food Price Inflation

Forecasting of any natural event is relatively difficult and even impossible in some instances, especially when we consider the limitations of our own minds when handling enormous amounts of quantitative data. For food prices, which are impacted by international and domestic factors, it is almost impossible for a single individual to comprehend the full scope of attributes that can influence the cost of a single product. In addition to location based factors, food prices are also affected by countless other forces including supply, demand and the value of the trading currency.

Traditionally, food forecasts have been performed by financial and reporting firms which speculate internally to form general assumptions and make predictions surrounding demand and supply. These predictions are then used as the basis for forecasting tools. The results are as would be expected, widely varying in predicted value and error rate (Joutz F. L., 1997). The Economic Research Services (ERS), an arm of the United States Department of Agriculture (USDA) is the only United States Government entity which regularly examines food prices to produce forecasts. In its own 2000 report on its forecasting processes from 1984 to 1997, the ERS revealed that the simple univariate techniques found in the menu-based time series forecasting system in SAS suite of analytics software produced a lower Root Mean Square Error (RMSE) in 7 of 23 price indexes and comparable RMSE rates in another 13 (Joutz, Trost, Hallahan, Clauson, & Denbaly, 2000). The report indicates that the longstanding ERS approach to forecasting through a process of crafting assumptions about supply and demand was only able to outperform the baseline models in 3 of the 23 price indices. To put it bluntly, the ERS despite its educated assumptions was outperformed by baseline statistical models which can be easily accessed by a non-expert forecaster.

1.1.4 Machine Learning Forecasts

The ability to precisely forecast price increases to circumvent inflation will become one of the most prioritized accomplishments for any country that continues to thrive in the 21st century. Governments, businesses and consumers engaged in forecasting prices and interpreting price information are likely to find support for their efforts in the domains of machine learning and statistical analysis as these methods of generating meaningful information have the potential to precisely predict inflation rates. Machine learning is a multifaceted discipline in which there are numerous proven techniques for identifying strategic information and solving complex problems. Statistical analysis has existed for over a century and is routinely used in business and academia to derive meaningful information from large stores of data.

In the realm of food price predictions, the combination of both disciplines could not only generate knowledge, which could potentially forecast important increases in costs, but also can frame newly discovered knowledge in a manner that can be understood with relative ease by all levels of society. These results could then be used to inform consumer budgets, maintain supermarket inventories or even influence federal trade policy. To finally overcome the struggles caused by the volatility of price inflation as it silently plagues our markets would require an extraordinary endeavor and this research thesis aims to deliver an illustration of that effort in the form of a machine learning approach to the prediction of consumer food prices.

Machine learning forecasts have become popular and even common place in financial and commodities markets (Kim, 2003) (Cao & Tay, 2003) (Ticlavilca, Feuz,, & McKee, 2010). This is due to a heighten recognition for the power of Machine Learning techniques in handling with complex tasks and nonlinear data. On the other hand, economics which is the field where food prices are usually examined, is widely regarded as a Humanities subject separate from the business and finance, was less impacted by this trend.

1.1.5 A Machine Learning Approach to Food Price Predictions

For decades, economists – including those at the ERS – have attempted to connect commodity prices and consumer prices (Akram, 2009), (Webb, 1988), (Cody & Mills, 1991) but as Blomberg and Harris explain, while this may have been true in the 1970's and 1980's, the same theory does not apply today. The food processing industry and the financial markets that support it have increased in complexity to such an extent that input commodity prices have little impact on the cost of the final product. Commodities simply do not hold the same importance that they once did and financial shocks at the commodity level have only a slight impact on the overall economy (Blomberg & Harris, 1995).

This research thesis with its machine learning approach is aimed at breaking the mold that constrains food price forecasts. By incorporating machine learning techniques for data analysis over the conventional approach of expert generated assumptions, the hope is to identify strong relationships and reveal more complex views of the marketplace to produce far more robust forecasts.

1.2 Research Objectives

This research thesis attempted to answer the following question, "What is the best way to predict food prices for the average Canadian consumer?" The overall objective was to forecast the Canada Consumer Price Index (CPI) to assess the performance of various machine learning techniques against three data models. Specifically, the research aimed to:

1. Determine the top performing model of the three models assessed
   i. Holt-Winters
   ii. Food Price Report
   iii. Financial Futures-based Markets
2. Determine the top performing machine learning technique of the four assessed
   i. Multivariate Regression: Linear Regression
   ii. Artificial Neural Networks: Multilayer Perceptron
   iii. Support Vector Machine: SMOreg
   iv. Decision Tree: M5P Tree
3. Evaluate the performance of the Multilayer Perceptron as the only technique to incorporate backpropagation.

This research thesis was not meant to be a definitive approach to food price forecasts by favoring one technique over another but rather was intended to illustrate the accuracy of machine learning techniques in this forecasting domain by using the tools and techniques which can easily be duplicated by the average consumer.

1.3 A Machine Learning Approach

The primary objective of this research thesis was to employ a machine learning approach to the creation of econometric models, which can then be used to forecast food prices in the major food categories as outlined by the Canada Consumer Price Index (CPI) and published by Statistics Canada. The research compared a benchmark model currently being used by financial analysts and economists to identify simple trends in time-series data, with two other models designed to use historical datasets as predictors.

1.3.1 Benchmark Model

The Holt-Winters triple smoothing exponential technique is quite popular in statistical and economic analysis for forecasting data points in a time series because it incorporates baseline, trend and seasonal analysis to forecast data points of seasonal data. Even though it is an older technique, originally developed in 1957, Holt-Winters has been proven to deliver satisfactory results. In 2016, Tratar and Strmcnik compared the Holt-Winters method to a multiple regression method for long-term and short-term heat load at an electricity generating company. The Holt-Winters method bested the multiple regression method in 4 of 6 experiments and completely dominated the long-term forecasting category (Tratar & Strmčnik, 2016).

1.3.2 Food Price Report Model

A dataset adapted from that used in the Food Price Report 2017 served as the data source of several machine learning models. The dataset consisted of historical data features in 20 groups of variables, listed below, identified by the report's authors as being correlated with the volatility of food prices and the Canadian consumer price index. More details on the features and sources of the dataset are available in Appendix A.

1. Canada Household Income
2. Canada Household Debt
3. Canada Immigration
4. Canada Income Distribution
5. Canada/US Unemployment
6. World Bank Commodity Prices
7. FAO Commodity Index
8. International Aid Dollars
9. World Population
10. World Diesel Prices
11. World Gasoline Prices
12. Crude Oil Prices
13. Canada CPI: Energy
14. World Agricultural Food Production
15. World Rainfall
16. World Temperatures
17. Canada-US Exchange Rate
18. US Credit to Agriculture
19. US Overnight Lending Rates
20. World Fertilizer Price

1.3.3 Financial Futures-Market based Model

Owing to its purpose, the financial futures-market inherently predicts commodity prices at the market level. This research thesis speculated that these predicted market prices could themselves be predictors of consumer prices. The financial futures-market based model is built on historical closing prices of commodities at seven international exchanges, listed below. The features and sources of the datasets are available in Appendix A.

1. Chicago Board of Trade (CBOT): Corn, Oats, Rough Rice, Soybeans, Soybean Meal, Soybean Oil, Wheat
2. Chicago Mercantile Exchange (CME): Class III Milk, Feeder Cattle, Live Cattle, Live Hogs

3. Intercontinental Exchange (ICE): Cocoa, Coffee “C”, Orange Juice, Sugar #11
4. Kansas City Board of Trade (KCBT): Wheat
5. London Commodity Exchange (LCE): London Cocoa, London Sugar
6. Minneapolis Grain Exchange (MGE): Wheat
7. Winnipeg Commodity Exchange (WCE): Canola

The best performing model and machine learning techniques were identified by their ability to achieve the lowest Mean Absolute Percentage Error (MAPE) below that of the Holt-Winters benchmark model. As a tool for hedging and speculation, futures trading provides exceptional convenience and economies of transactions allowing it to possess similar qualities to that of spot markets and is consequently ideal for forecasting prices (Working, 1976). This researcher hypothesized that the financial futures-market based model would outperform the other models because accurate price predictions are an inherent and critical component of trading in the financial futures-market.

The remainder of this research paper is organized into chapters, starting with a background and related work, followed by the research methodology and a description of the experiment evaluation. The results are revealed at the end and conclusions were drawn on this basis of the results achieved

CHAPTER 2: BACKGROUND AND RELATED WORK

2.1 The Global Context

Agricultural and food prices have been on the rise since the mid-2000s and have now become parallel with the prices of other commodities (European Union, 2015). During this period, consumers in both developed and developing countries have experienced the hardships caused by higher food prices and increasing input costs, perpetuated by a cycle of rising prices. The most obvious driver of the price spike in the previous decade has been the worldwide increase in demand. Since World War II and after the introduction of the Baby Boomer generation the world's population has tripled in size and with the average person on earth generally living a longer life, under much better conditions than those before World War II, we have been able to consume more food, products and services (European Union, 2015).

In addition to the natural improvements to a country's food supply and the associated improvement of the nutritional state of its population, the promotion of healthy lifestyles also reduces the burden of incommunicable diseases, increases the productivity of human capital and ultimately leads to a better quality of life (Food and Agriculture Organization, 2002). Thus, since the last World War, many countries have strived to incorporate some form of "nutrition transition" and have been supported in finances and efforts by a variety of international aid organizations. Apart from countries in economic transition, such as those recovering from civil war, natural disaster or other forms of national emergency, developed and developing countries have all together found a way to maintain a gradual increase in their daily calorie consumptions per person from both plant-based and animal-based sources (Food and Agriculture Organization, 2002).

The increased demand for food from both developed and developing countries has caused food commodity prices to skyrocket on the international market. In the domestic markets, increases in income and improvements to the accessibility of foreign food markets have enhanced the demand for a variety of foods, which in turn has caused prices to increase. These increases have since trickled down from the commodities to the consumers. Between 2005 and 2007, just prior to the Unites States housing crisis and subsequent global recession, the prices of wheat, coarse grains, rice and oilseed crops all nearly doubled. Meanwhile, from 2005 to 2008 the Food and Agricultural Organization's (FAO) food price index indicates an 83% increase in global food prices at the consumer level. This increase was almost three times (3x) higher than the previous 3 years, 2002 to 2005, and almost three times (3x) higher than the following 3 years, 2008 to 2011 (Food and Agriculture Organization, 2017).

The causes of this spike in prices are complex and due to a combination of mutually reinforcing factors (Organization of Economic Co-Operation and Development, 2008) including:

- weather and climate related events in key grain-producing regions that caused negative yields
- low stocks for cereals and oilseeds caused by increased demand in developing countries
- increased feedstock use in the production of biofuels owing to policy-driven demand

- increased production costs brought on by rapidly rising oil prices and price jumps due to a continuing devaluation of the US dollar as the currency in which indicator prices for these commodities are typically quoted.

Production yields and macroeconomic conditions have also had a significant impact on food prices. The negative yields, such as those observed in 2015 – 2017 due to unusual weather conditions in California (Austen, 2016), may have been temporary but growing concerns for the impact of long-term climate change and pollution may prove to be well-founded as environmentalists continue to forecast more permanent effects which will undoubtedly also impact food prices (Bandaraa & Caib, 2014), (Wheeler & Braun, 2013).

Meanwhile, the macroeconomic conditions that favor economic growth, increases in purchasing power, and stronger demand for agricultural commodities are permanent factors in the determination of food prices. Additionally, the demand for food is less responsive to changes in prices when the purchasing power of the consumer increases (Öner, 2012). So, with each country actively striving for economic growth it is not difficult to see how in the absence of reactionary demand food prices can increase to inflationary rates due to growing consumer incomes in many developing countries. Consumers of food products eventually lose some of their purchasing power even when their income increases because the higher price becomes the standard for all products and services. In the context of the financial world, this type of non-participatory price increase in food would be equivalent to an impossible situation where stock prices could be manually manipulated by the listed companies without consequence and without the involvement of the market who would then be expected to honor the new prices.

The thinness of markets, or the share of imports and exports relative to the size of global consumption or production, can have large scale effects on global prices because of protectionist federal policies and imperfections in domestic markets (Organization of Economic Co-Operation and Development, 2008). In other words, increased demand for a particular food item in one country, such as corn for biofuels in the United States, could have the effect of price increases in the US domestic market which are not aligned with the price rates of international trade. Thus, world market prices would adjust to accommodate an external shock to demand and traded quantities. In the case of biofuels, policy-driven demand for biofuel production is generally considered less responsive to prices than traditional food demand and is a strong factor in the increase of fuel and animal feed prices (Organization of Economic Co-Operation and Development, 2008).

2.2 Canada, the Food Price Report

Originally named the Food Price Index (FPI), The Canada Food Price Report was created at University of Guelph by Dr. Sylvain Charlebois and Dr. Francis Tapon. Since 2010, the Food Price Report has been published as a tool that focuses on factors affecting the future of consumer food prices over a 12-month period (Wilson, 2015). The report is now published by Dalhousie University and includes a machine learning methodology which is further supplemented by the expert advice of its prominent authors.

The 2017 report features a range of models from simple regression to neural networks and twenty independent variables, identified as potential inputs due to high correlations with a specific food category of the consumer price index. Also, included in the report is an interpretation of economic drivers, fundamental to the movement of food prices in Canada.

These economic drivers are divided into three main categories, macro drivers such as energy cost and climate change, sectorial drivers such as the food processing industries and consumer food awareness, and domestic drivers, such as consumer debt, consumer income and income distribution.

The authors' understanding of these drivers played a critical role in the report's forecast and each group of drivers is reviewed yearly to reflect the potential impact on food prices. The groups of drivers formed the basis for the twenty independent variables utilized in the 2017 report. It is helpful to consider how the size and impact of these groups of drivers has changed over the years to maintain alignment with the current state of affairs. While the number of macro and sectorial drivers has slightly increased in the past five years, the number of domestic drivers has decreased. Although the authors offer no specific explanation for the removal of the domestic driver, during the report's history the local influences on price had taken a back seat to international pressures. Given that Canada imports most of its food, roughly 80% of it being fresh fruit and vegetables, it is not difficult to see how domestic drivers could be eliminated and replaced by more impactful global drivers.

In this research thesis, the feature selection process played a role like that of the report's authors. Both attempted to determine the impact of each group of variables then used this knowledge to build a predictive model. It was not clear if the machine learning models in this research thesis would up hold the conclusions of the report's authors regarding the importance of the macro-economic drivers. Unfortunately, such a comparison was outside the scope of the research thesis and was not included in the analysis of the model.

Nevertheless, the similarities between the feature selection process and the authors' views on the changing macro, sectorial and domestic drivers should not be ignored.

2.3 A Review of Related Work

2.3.1 Forecasting with Futures-Market Data

The debate over futures markets as agencies for rational price information is by no means new, Tomek in 1997 famously concluded that futures prices can be viewed as forecasts of maturity-month prices, and the evidence suggests that it is difficult for structural or time series econometric models to improve on futures forecasts (Tomek, 1997). In 2008, this conclusion was upheld by Carter and Mohapatra with respect to the non-storable commodity, hog futures. They examined hog futures prices from 1994 to 2008 and determine via statistical means that the hog futures market was an unbiased predictor of cash prices (Carter & Mohapatra, 2008). Similarly, Bowman and Husain in 2004, assessed the performance of three types of commodity price forecasts based on judgement, historical data, and commodity futures. They determined, in the long term, spot prices tended to move towards futures prices and exploiting this knowledge produced much better forecasts than the other two forecast types. They concluded based on statistical and directional accuracy, commodity futures yield the best forecasts over long horizons (Bowman & Husain, 2004).

2.3.2 Artificial Neural Networks

Adequate research and a clear consensus around the predictive value of futures-market prices already exist in the literature and the techniques associated with these predictions have similarly been vigorously studied. Specifically, Artificial Neural Networks (ANN) have increased in popularity in recent years, being used noticeably in financial applications. Researchers have recognized the predictive value of the ANN and employed it to tackle various real world problems including bankruptcy predictions (Guoqiang Zhanga, 1999), forecasting exchange rates (T.H. Hann, 1996) and predicting the stock market (de Faria, Albuquerque, Gonzalez, Cavalcante, & Albuquerque, 2009). In 1996 Chiang et al. forecasted the end of year net asset value (NAV) of mutual funds using an ANN and traditional econometric techniques. The ANN achieved a Mean Absolute Percentage Error (MAPE) 42% lower than the closest traditional technique. In fact, it bested the traditional techniques in 4 of the 5 mutual fund categories surveyed (Chiang, Urban, & Baldridge, 1996).

2.3.3 Forecasting Food Prices with Machine Learning

Machine learning approaches to predicting food prices are less popular in the literature and was the focus of this research thesis to extend the work in this domain. Nowrouz Kohzadi et al., (1993), used monthly live cattle and wheat prices from 1950 to 1990 to compare an Autoregressive Integrated Moving Average (ARIMA) technique to an ANN. The artificial neural network model achieved a 21% and 26% lower MAPE respectively (Nowrouz Kohzadi, 1996). More recent attempts have achieved even more outstanding results. In 2017, Malhotra and Maloo conducted a machine learning approach to forecasting food inflation in India. Their experiments yielded an R-squared value of 99.1% from 25 instances of data (Malhotra & Maloo, 2017). Another recent study from Zou Haofei et al. (2007) focuses on backpropagation as a means of optimizing the performance of the neural network. Backpropagation appears throughout the literature as a valuable technique for forecasting time series data. Haofei et al. (2007) applied a multistage approach to the conventional backpropagation algorithm to forecast the price of wheat on the China Zhengzhou Grain Wholesale Market. The results indicate that while the conventional backpropagation algorithm may have struggled against the baseline ARIMA model in the in-sample tests, both the conventional and multistage approaches significantly outperformed the benchmark model with MAPEs 36% and 64% lower respectively (Haofei, Guoping, Fangting, & Han, 2007).

Unfortunately, for this research thesis a comparative number of training instances, as is present in Haofei et al. (2007) study was not available for performing a similar multistage backpropagation approach. However, forecasting with limited data instances to forecast has been shown to deliver satisfactory results, owing to improved techniques. Wolczak in 2001 presented evidence contradicting the current financial neural network development heuristic, which implies that greater quantities of training data produce better-quality forecasting models. He showed that given a small amount of historical knowledge, around 1 or 2 years, a neural network can forecast future currency exchange rates with 60 percent accuracy. While neural networks trained on a larger training set have a worse forecasting performance.

The literature shows that the use of machine learning techniques, including ANNs and backpropagation, as a method for predicting time series, delivers interesting results and is a promising alternative to traditional approaches (Zhang, Patuwo, & Hu, 1998). Additionally, futures-market data has been shown to have a side benefit of producing good price forecasts (Emmons & Yeager, 2002). The combination of these techniques into a single data model was likely to yield superior results.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Experiment Overview

1. Three datasets were exposed to various forecasting techniques, with the intended purpose of forecasting each major food category of the Canada Consumer Price Index to create three data models and determine which model produces the lowest error rate.
2. The first dataset, the Canada Consumer Price Index, was tested for seasonality then became the input layer for the Holt-Winter Triple Smoothing technique. Each major food category of the CPI was exposed to the Holt-Winters technique to forecast its value in 2016 and the resulting data model served as the benchmark model for the remaining data models.
3. The second dataset, the Food Price Report data, was firstly reduced by the process of Correlation-based Feature Selection to create a unique dataset for each major food category of the CPI. The reduced unique datasets then became the input layers for various machine learning techniques to forecast the corresponding CPI values in 2016 and create the Food Price Report model. The lowest error rate per food category of the new model and the techniques that achieved those rates were recorded for comparison against the benchmark model and the final data model.
4. The third dataset, the Financial Futures-Market data, was also reduced by the process of Correlation-based Feature Selection to create a unique dataset for each major food category of the CPI. The reduced unique datasets then became the input layers for various machine learning techniques to forecast the corresponding CPI values in

2016 and create the Financial Futures-Market model. The lowest error rate per food category of the new model and the techniques that achieved those rates were recorded for comparison against the benchmark model and the previous Food Price Report model.

5. The resulting error rates in each food category were finally compared across the data models to determine which model produced the lowest average error rate.
6. The model with the lowest average error rate was then subjected to further analysis in the form of backtesting to determine if the model's predictive value extends to its historical data.
7. If the backtesting analysis yielded positive results, then the combination of forecast analysis in the form of the lowest error rate model and backtesting analysis in the form of the backpropagation algorithm would yield a higher predictive value than the other techniques tested with that model. Further analysis of the Multilayer Perceptron artificial neural network and its backpropagation algorithm on the model with the lowest average error rate was performed to determine the validity of this claim.

3.2 Dataset: Canada Consumer Price Index

The primary goal of the machine learning model is to forecast, for the coming year, the Canadian Consumer Price Index, which is published each month by Statistics Canada. The consumer price index was chosen for this task because it is the de facto measure of inflation in Canada and is in fact used extensively by the Bank of Canada to chart inflation and adjust interest rates. In this way, the Consumer Price Index becomes the dependent variable as well as the final output produced by the model. Eight categories of the consumer price index are examined during the process:

1. Meat: Fresh or frozen beef, pork, poultry and other processed meats such as ham and bacon, excluding seafood.
2. Seafood: Fresh, frozen, canned or otherwise preserved seafood including fish and other marine products.
3. Dairy: Fresh milk, butter, cheese, ice cream and other related processed products, also includes eggs.
4. Bakery: Breads, cookies, crackers and other bakery products along with rice, pasta, flour and cereal products excluding infant formulas.
5. Fruit: Fresh and preserved fruit including apples, oranges and bananas along with fruit juices and nuts.
6. Vegetables: Fresh, frozen, dried, canned or otherwise preserved vegetables including potatoes, tomatoes and lettuce.
7. Other: Sugar and confectionery products, edible fats and oils, coffee and tea, condiments, spices, vinegars, soups, infant formula, pre-cooked and frozen food preparations, non-alcoholic beverages.
8. Food Purchased from Restaurants: Table-service, fast food, take-out, cafeterias and other restaurants.

Ultimately, three models were developed to forecast each food category of the consumer price index by using various unique datasets of publicly available historical records. The Consumer Price Index was the actual value being predicted by the models and the accuracy of these predictions was used to evaluate the performance of each model to further determine the best performing model per food category.

3.3 Holt-Winters Triple Exponential Smoothing

The first step in the process of determining the best performing model per food category was to create a benchmark calculation, which was used to judge the performance of the other computer generated models. The Holt-Winter Triple Exponential Smoothing technique represents a popular form of time-series analysis used in statistics and econometrics that has been peer reviewed and proven to deliver satisfactory results. The method comprises of a forecasting equation $\hat{y}_T$ along with three smoothing equations for level a_t, trend b_t and seasonality s_t, hence the triple smoothing reference in its name. There are also three smoothing parameters α, β and, γ along with a period parameter p and the initial value is given as Y. The multiplicative form of Holt-Winters Triple Exponential Smoothing model (Hyndman & Athanasopoulos, 2013), as shown below, was used in this experiment.

Level: $$\mathrm{a}_t = \alpha \frac{Y_t}{\mathrm{s}_{t-p}} + (1-\alpha)(\mathrm{a}_{t-1} + \mathrm{b}_{t-1})$$

Trend: $$\mathrm{b}_t = \beta(\mathrm{a}_t - \mathrm{a}_{t-1}) + (1-\beta)\mathrm{b}_{t-1}$$

Season: $$\mathrm{s}_t = \gamma \frac{Y_t}{\mathrm{a}_t} + (1-\gamma)\mathrm{s}_{t-p}$$

Forecast for time period $T+\tau$: $$\hat{y}_{T+\tau} = (\mathrm{a}_T + \tau\ \mathrm{b}_T)\mathrm{s}_T$$

Given its track record and robustness, the Holt-Winters method was selected as the measuring stick for the more data intensive food price report model and financial futures-market based model.

The Holt-Winter method is specifically designed to cope with trends and forecast seasonal data, i.e. data points that are repetitive over some period. The triple smoothing method is useless on non-seasonal data, so to determine the presence of seasonality in the CPI a stationarity test was performed on each food category of the index using a the XLSTAT software, version 19.02.44125. Per XLSTAT, a time series is said to be stationary if its statistical properties such as mean, variance and autocorrelation do not vary with time. Simply, the stationarity tests determined whether the properties of the consumer price index datasets depend on the time at which the time series is observed. If a dataset is stationary, then it lacks the trends and seasonality which would allow it to be effectively modelled by the Holt-Winters Triple Smoothing technique, (Hyndman & Athanasopoulos, 2013). The results of the stationarity tests are available in Appendix B.

Per the XLSTAT website, the software is a suite of statistical add-ins for Microsoft Excel that has been developed since 1993 by Addinsoft to enhance the analytical capabilities of Microsoft Excel. Since 2003, Addinsoft has been a Microsoft partner and all the XLSTAT analytical add-ins are registered on the Microsoft Office Marketplace. The XLSTAT software relies on Microsoft Excel for the input of data and to display results. This makes the software very convenient for exporting results. Computations are done using autonomous software components that are optimized for speed and efficiency. The XLSTAT results were benchmarked against other statistical packages to support the integrity of its results.

The Holt-Winters Triple Exponential Smoothing technique was accessed via the Waikato Environment Knowledge Analysis (WEKA), version 3.8.0. WEKA is the product of the

University of Waikato in New Zealand and was first implemented in its modern form in 1999 and is licensed under the GNU General Public License (GPL). WEKA is a data mining workbench and consists of a collection of machine learning techniques for data mining tasks. The WEKA suite contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization as well as several add-on packages for specialized tasks in areas such as time-series analysis. It incorporates several popular machine learning techniques into a streamlined user interface which is widely used in academics. These advantages make WEKA well-suited for developing new machine learning strategies. Most notably, WEKA is used in the classroom as a teaching tool and in research to first verify whether a hypothesis has a high likelihood of success by quickly creating an initial model which can then be expanded by more sophisticated tools such as R. For this research thesis the WEKA Time Series Package, part of a new release of dozens of data mining packages available since version 3.7.3, was used to provide all the analytical processing required to determine the best performing model per food category.

3.4 Dataset: Canada Food Price Report

The food price report model was based on an identical dataset that was used during the creation of the Canada Food Price Report 2017. The initial dataset, before the correlation-based feature selection process, consisted of 476 features, spanning 17 years, from 1999 to 2015. The decision to cap the historical data at 1999 was two-fold, firstly some selected data sources were unable to supply reliable data prior to 1999, secondly the Canada Harmonized Sales Tax, implemented in 1997 by the Atlantic provinces, has had the effect of lowering consumer prices and making comparisons to the period prior more susceptible to error (Smart, 2011).

The dataset was ordered chronologically from oldest to most recent and the model was trained on 80% of the data and tested on the remaining 20% at a 95% confidence interval. Further details about the data and data sources are outline in Appendix A. Correlation-Based Features Selection (CFS) was not incorporated in the original Food Price Report 2017 methodology however it was applied in this research thesis. The CFS procedure is a statistical method of determining which features are most likely to impact the individual food groups represented by the consumer price index. The results of the CFS process are outlined in Appendix E.

The historical records were used to create a general assumption of how these features have impacted food prices in the past and what could be their likely impact in the future. The CFS process identifies features which are highly correlated with each consumer price index food category (the dependent variable) but are not correlated with other features in the dataset. Historical records pertaining to each identified feature were grouped together and formed a specialized dataset for each food category of the consumer price index. These specialized datasets then served as the basis for the computer-generated models created by several well-known techniques. A range of regression techniques were selected to offer various approaches to forecasting the consumer prices index. Four techniques were selected from among the WEKA offerings:

- Multivariate Regression: Linear Regression
- Neural Network: Multilayer Perceptron
- Support Vector Machine: SMOreg
- Decision Tree: MP5

3.5 Dataset: Financial Futures-Market

This research thesis aimed to investigate the value of the financial futures-market as a predictor of consumer food prices. The financial futures-market model formed one of three models which were compared to determine the best method for forecasting food prices. The initial dataset, before the correlation-based feature selection process, consisted of 20 features, spanning 16 years, from 2001 to 2015. The decision to cap the historical data at 2001 was based on the availability of data for the Moore Research Center. Each feature represents the annual average end-of-day prices for various food commodities at several global commodities markets as extracted from the Moore Research Center website. The dataset was ordered chronologically from oldest to most recent and the model was trained on 80% of the data and tested on the remaining 20% at a 95% confidence interval. Further details about the dataset are outlined in Appendix A. This model also applied the Correlation-Based Features Selection (CFS) process and the results are outlined in Appendix E.

3.6 Correlation-based Feature Selection (CFS)

Feature selection is the process of identifying and removing as much irrelevant and redundant data from a dataset as possible. By reducing the dimensionality of the dataset, the size of the hypothesis space is also reduced, allowing algorithms to operate faster and more effectively; in some cases, the accuracy of forecasts can also be improved, (Hall M. A., 2000). Feature selection is of great importance to the domains of data mining and machine learning. As the amount of available data continues to grow exponentially, the difficulty of testing and training general classification models also grows due to the increase in the dimensionality of the data.

Karegowda, Manjunath and Jayaram in 2010, compared feature sets determined by CFS to those determined by Information Gain. The correlation based feature set improved upon the gain ratio feature set, confirming CFS as a successful approach, even more advantageous than information gain, as a sophisticated tool for data analysis (Karegowda, Manjunath, & Jayaram, 2010). Another relevant paper comes from Yu and Liu in 2003, who introduced a novel approach to CFS with the Fast Correlation-based Feature selection process which as their results indicated, greatly outperformed other methods in real world comparisons with respect to speed and feature reduction. Considering the results from Karegowda et al. and Yu et al. the CFS method was selected to serve as the dimensionality reduction technique.

This research thesis presented a relatively new approach to feature selection in Correlation-based Feature Selection (CFS) to determine the merit and rank of each dataset's features. The central hypothesis behind the CFS is, good feature sets contain features that are highly correlated with the class yet uncorrelated with each other (Hall M. A., 1999). The CFS technique utilized a heuristic approach by evaluating the merit of subsets of features instead of individual feature instances. The usefulness of individual features is accounted for by incorporating the overall feature strength of predicting the dependent variable and the level of correlation among those features. The CFS process examined many feature subsets, if the dataset had 'n' possible features in a dataset then there could be '2^n' possible subsets to be examined and the only way to find the best subset would be to evaluate them all, (Hall M. A., 2000).

3.7 Mean Absolute Percentage Error

With the Holt-Winters model as the benchmark calculation, the Food Price Report and Financial Futures-Market based models were evaluated against it to determine if either produces a lower error rate. This evaluation was performed by using the Mean Absolute Percentage Error (MAPE) metric, which is outputted by the WEKA workbench. The MAPE was calculated for each food category of the consumer price index in both the Food Price Report model and the Financial Futures-Market based model.

The Mean Absolute Percentage Error (MAPE) is a popular measure of performance in forecasting. Two of the major reasons for its popularity are its ease of interpretation and its scale independency, which allows it to be used across different datasets. In some forecasting problems, the MAPE may be a good alternative to the popular Mean Squared Error (MSE) (Lam, Mui, & Yuen, 2001). However, many authors have also argued against the use of the MAPE in certain situations. MAPE measures have the disadvantage of being infinite if the values of the dependent variable are equal to zero for any period and having an extremely skewed distribution if any value is even close to zero (Hyndman & Koehler, 2006).

For this research thesis, the disadvantages associated with the MAPE were minimal as the dependent variable, the Consumer Price Index (CPI) per major food category, had no values at zero or near zero. Despite the varying views on the MAPE metric it was selected as the measure of comparison for this research because of its popularity in finance and economics, as evidenced by its availability in the analytical software WEKA, which was utilized in this research thesis (Lam, Mui, & Yuen, 2001). The average MAPE error was calculated for each food category in each model (Holt-Winters, Food Price Report and Financial futures-market based) and the model with the lowest average error was deemed the top performing model.

3.8 Backtesting

Backtesting is the process of applying a predefined model to historical data to test the accuracy of the model in the past, this provides an additional layer of analysis for evaluating the performance of the model. In backtesting a predefined model is applied to historical records and the forecasted value per historical record is then compared to the actual record to determine the error generated by the model. The error rate across the historical records is helpful in determining the predictive value of the model. The combination of backtesting analysis along with forecast analysis, as explained below in Chapter 3.10 Backpropagation, allow for the model to be evaluated in the past as well as in the future.

In this research thesis backtesting analysis was performed on the model with the lowest average error rate. The dataset was ordered chronologically from oldest to most recent with 30% of the dataset being used as input for the model, 50% of the records were tested to evaluate the model's historic predictive value through backtesting and the final 20% of the records in the dataset were tested against the model, as trained on the previous 80%. The errors were visualized to aid in the analysis by highlighting the trend of the errors produced over time. In a perfect world, the error produced by the model per record would have decreased or remained consistent over time. The only drawback to backtesting is that the error results are not usually incorporated into the model to optimize its performance.

3.9 Backpropagation

Backpropagation is one of the most widely applied neural network architectures today after being introduced by Bryon and Ho in 1969, however it was not made famous until 1980 by Rumelhart, Hinton and Williams. Its popularity primarily revolves around the ability of backpropagation networks to learn complicated multidimensional mapping (Hecht-Nielsen, 1988). Through backpropagation, networks can be trained to identify previously observed characteristics in inputs, then create outputs based on whether a new input exhibits similar characteristics.

To incorporate this advanced learning architecture of backpropagation, the Multilayer Perceptron technique found in the WEKA workbench was selected as the neural network model in the food price report model as well as the financial futures-market model. The technique trained on the data for 500 epochs, with hidden layers equal to (the number of features + 1) / 2, no decay rate, a momentum of 0.2 and a learning rate of 0.3. The Multilayer Perceptron's backpropagation algorithm allowed for the back-testing process to occur until the set time frame of 500 epochs elapsed, which means the food categories could have experienced varying degrees of backpropagation during the experiment. WEKA also provides a graphical interface for altering the neural network while the model is still training but this option was not utilized during the process.

Backpropagation is not available with the other techniques used in this study as it is primarily a feature of neural networks. Backpropagation provides a much more robust method of analysis and the Multilayer Perceptron model was hypothesized to outperform the other machine learning models.

CHAPTER 4: EXPERIMENTAL EVALUATION

4.1 Holt-Winters Triple Exponential Smoothing

The Holt-Winters Triple Exponential Smoothing model is a popular time series technique in econometrics. Dating back to 1957, the technique is attributed to Charles Holt, a professor working at Carnegie Mellon University, who described the original method which was later extended by his student Peter Winters in 1960. The method comprises of a forecasting equation along with three smoothing equations for level, trend and seasonality, hence the triple smoothing reference in its name.

The Holt-Winters method is made available in the WEKA Time Series package. The WEKA documentation does not indicate whether a multiplicative or additive approach is taken and does not provide this selection as an option; however, given that the Time Series package automatically checks for seasonality before proceeding with a forecast, the assumption was made that the package also determines to correct approach. The scope of this research thesis is such that the specific method used had only a small impact on the outcome. This research thesis is not meant to be a definitive approach to food price forecasts, favoring one technique over another but rather was intended to illustrate the accuracy of machine learning techniques in this area of forecasting by using the tools and techniques which can easily be duplicated by the average consumer.

4.2 Mean Absolute Percentage Error

To compare models across different datasets requires a scale independent metric. Percentage errors provide this advantage and are frequently used to compare performance in such situations. One of the most common percentage errors is the Mean Absolute Percentage Error (MAPE). MAPE is readily available in the WEKA workbench, making it an obvious choice.

Percentage errors do have the disadvantage of being infinite if the values of the dependent variables are equal to 0 for any period of interest and having extreme values if the dependent variable values are even close to zero. Thankfully this is not the case with the CPI values in this dataset. Percentage errors such as the MAPE also assume a meaningful zero which the CPI values do provide in that a 0 CPI indicates that price levels are as there were when the study first began (Hyndman & Athanasopoulos, 2013). Hyndman and Koehler recommend the used of the MAPE over other scale independent metric if all data are positive and much greater than zero, as is the case with the CPI data used in this research thesis (Hyndman & Koehler, 2006). The MAPE was selected as the measure of comparison for the determining accuracy of the three models.

4.3 Research Objectives

4.3.1 Top Performances

A key outcome of this research thesis was to determine the top performing model and machine learning technique in the experiment. Two measures were used to evaluate the performances of the models. First, to provide a simple tally of the model performances, the number of categories in which a model achieved the lowest MAPE - irrespective of the machine learning technique - were counted. Next, to assign and exact predictive value, the average MAPE per model which is the average to the lowest MAPEs achieved in each category - irrespective of the machine learning technique – were calculated. The top model was determined by evaluating these measures to identify any outstanding performances.

Next, the top performing model was further evaluated with two measures, like those mentioned above, to identify the top performing machine learning technique. First, to provide a simple tally of the techniques performances, the number of categories in which a technique achieved the lowest MAPE were counted. Next, to assign and exact predictive value, the average MAPE per machine learning technique was calculated. The top technique was determined by evaluating these measures to identify any outstanding performances.

4.3.2 Backpropagation

Backpropagation is a machine learning procedure dating back several decades, which repeatedly adjusts the weights of the connections in the network to minimize a measure of the difference between the actual output vector of the network and the desired output vector as based on the changes in the historical data. Because of the weight adjustments, internal 'hidden' layers which are not part of the input or output come to represent important features of the model, and the regularities in the task are captured by the interactions of these units (Rumelhart, Hinton, & Williams, 1986).

The goal of backpropagation is to optimize the weights so the neural network, the multilayer perceptron model, can learn how to correctly map inputs to outputs by identifying previously observed characteristics in inputs to learn complicated multidimensional mapping schemes, then create outputs based on whether a new input exhibits similar characteristics (Hecht-Nielsen, 1988). The use of backpropagation in this research thesis was aimed at maximizing the prediction value of the model and to incorporate a technique which is growing in popularity in the forecasting literature. To do so, the Multilayer Perceptron technique was selected because it includes a backpropagation algorithm. It was hypothesized that the Multilayer Perceptron technique would outperform the other techniques in the experiment because of its backpropagation algorithm.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Model Performance

The analysis and predictions of the research thesis are intended to influence decision makers in industry, government and the home. Ultimately, the goal would be for this type of analysis to impact various policy decisions including monthly shopping budgets, federal trade policy, and super market prices. Decision makers should be able to use this type of analysis for offseting market trends to allow for growth and sustainability.

The results of the stationarity tests (Appendix C) indicated that each CPI food category has a significant level of seasonality and could be analyzed with the Holt-Winters Triple Smoothing technique. As hypothesized, the Financial Futures-Market based model outperformed the Food Price Report model with 1.6% and 2.4% average error rates respectively. Each model captured 4 of the 8 CPI food categories and apart from the CPI Seafood category in the case of the Food Price Report model, both models easily bested the popular Holt-Winters benchmark model, which produced an average error rate over three times higher than the machine learning models. The CPI Restaurant category, which was regarded in the literature as the most difficult to forecast because it is a composite good that contains several items (Joutz F. L., 1997), achieved the lowest error rate in both the Food Price Report and Financial Futures-Market models at 0.5031 and 0.5451 respectively. The CPI Meat category was the most difficult to predict with a 7.5% average error rating across all three models.

The results of the lowest Mean Absolute Percentage Error per model per CPI food Category are tabled below.

CPI Food Category	Holt-Winters	Food Price Report	Financial Futures-Market
Meat	11.7	3.3	*2.6
Seafood	5.5	6.2	*2.4
Dairy	7.5	3.3	*0.9
Bakery	11.0	3.0	*1.0
Fruit	7.8	*0.9	1.3
Vegetables	7.3	*1.1	2.0
Other	5.6	*1.2	2.3
Restaurants	8.0	*0.5	0.5
Average MAPE	8.0	2.4	*1.6

Table 1: Comparison of MAPE results from each model. (* Lowest MAPE per CPI food category).

Across all three models, the CPI Other category produced the lowest average forecast error. It was also 1 of only 4 CPI food categories to achieve an average error rating below 5% across all three models, the others included Fruit, Vegetable and Restaurants. These four categories were coincidentally bested by the Food Price Report model, which included several more domestically related features than the Financial Futures model, even though these 4 categories do not compromise of foods which are produced domestically in great quantities. This may hint to a greater difficulty in forecasting the price of foods which are predominantly produced domestically, namely the Meat, Seafood, Dairy and Bakery categories. It may also hint at a lack of meaningful domestic data for predicting domestic goods. In the absence of meaningful domestic data, the Financial Futures Model bested the Food Price Report Model in accurately forecast the domestically produced categories

When this research thesis began the CPI 2016 values were not available but since then the values have been released by Statistics Canada. In the table below there are the percentage error between the actual 2016 CPI values and the predicted 2016 CPI values per model per CPI food category.

CPI Food Category	Holt-Winters	Food Price Report	Financial Futures-Market
Meat	12.5%	3.8%	*2.3%
Seafood	18.2%	10.2%	*4.7%
Dairy	5.2%	6.7%	*2.7%
Bakery	7.1%	8.0%	*1.3%
Fruit	14.3%	*0.8%	1.1%
Vegetables	22.7%	*1.8%	3.7%
Other	12.8%	*2.0%	4.5%
Restaurants	9.6%	*1.3%	2.3%
Average Percentage Error	12.8%	4.3%	*2.8%

Table 2: Percentage change between the 2016 CPI predicted values from Table 1 and the 2016 CPI actual values (* Lowest error per food category).

The CPI actual to CPI predicted table closely mirrored the results of the lowest MAPE per model, per CPI food category in Table 1 above. Meat, Seafood, Dairy and Bakery categories were all bested by the Financial Futures-Market based model, while the Fruit, Vegetables, Other and Restaurant categories were each won by the Food Price Report Model. This again hints to a deficiency of domestic data as it is reasonable to believe that the categories which are produce domestically – Meat, Seafood, Dairy and Bakery – would have been bested by the Food Price Report model since it contains most of the domestic data.

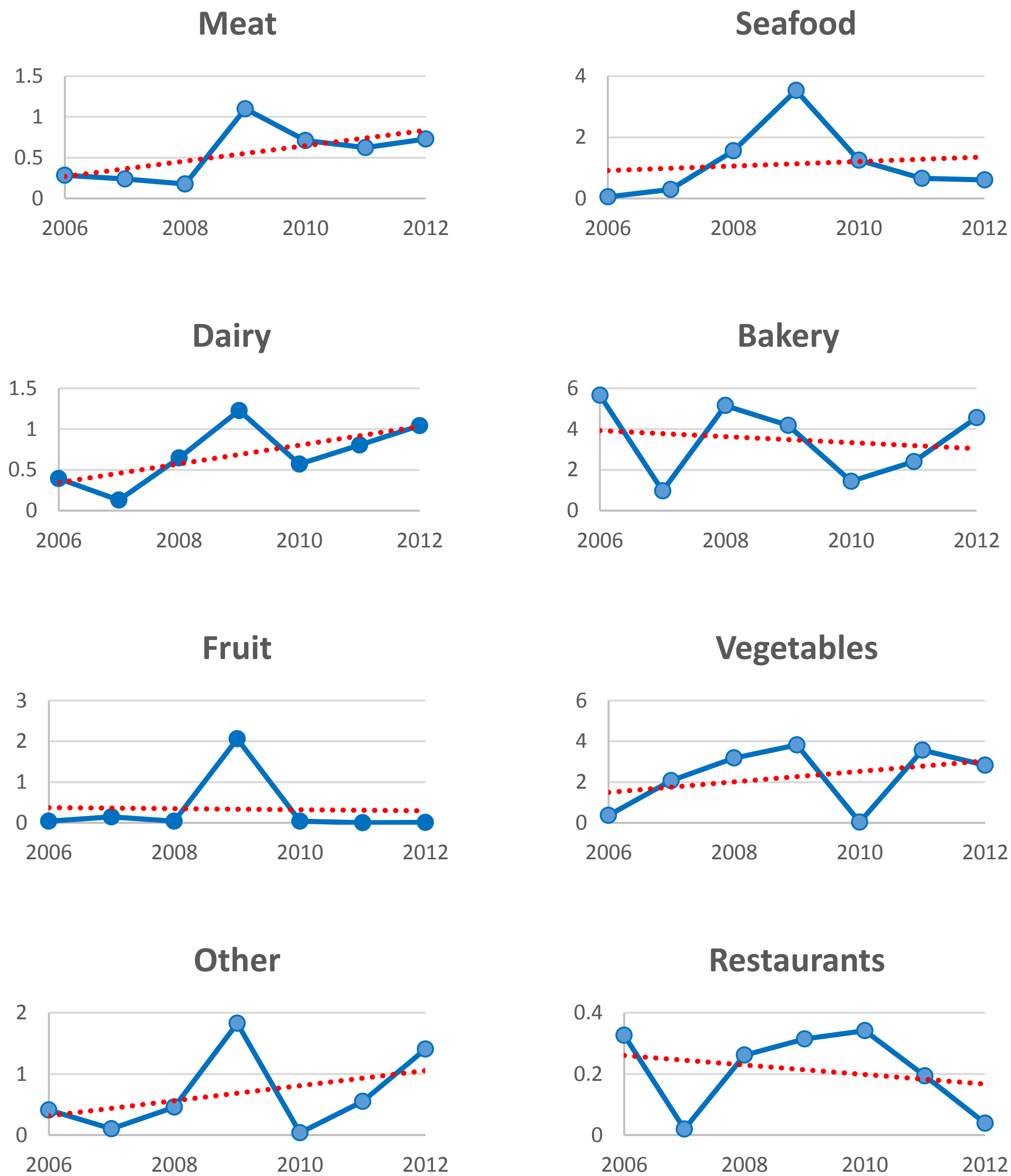
Overall the predictions were extremely close with 5 of the 8 categories having achieved error rates under 2% and only 1 category with a lowest error rate above 3%. The Financial Future-Market based model managed an average error rate of 2.8%, 1.5% below the Food Price Report model and 10% below the benchmark model, to solidify its position as the top performing model. The data from the Financial Futures-Market model, as the top performing model were then used to evaluate the top performing the technique and the results are resented in the following section.

5.2 Machine Learning Technique Performance

The backtesting results from the top performing model, the Financial Future-Market model, are visualized below in figure 2. The diagram illustrates the difference between the actual CPI values and the CPI values predicted by the model. A linear trend line has been inserted with Microsoft Excel to further illustrate the overall trend of the errors. If the backtesting analysis was successful then the trend lines would be slanted downwards to indicate reduced errors over time. Upward slanted trend lines indicate that the errors have increased over time.

Of the 8 CPI food categories, Bakery, Fruit and Restaurants are the only categories to have displayed a downward trend in errors during the period. Not surprising these were also the categories with the lowest percent error of 2016 actual values to 2016 predicted values. The underwhelming results of the backtesting analysis could indicate the need for more data as a lack of sufficient data points could have hurt the performance of the model.

Figure 1: Difference between actual CPI values and predicted CPI values in the Financial Futures-Market based Model.



None-the-less, the Multilayer Perceptron and its backpropagation algorithm still managed to best the other techniques in the experiment by having dominated in 5 of the 8 CPI food categories in the top performing model, the Financial Futures-Market as illustrated below in table 3. The Multilayer Perceptron, in addition to besting 5 of 8 categories in the Financial

Futures-Market model as noted previously and illustrated in table 3, also won 4 of 8 categories in the Food Price Report model for a combined top performance twice as good as the other techniques. The Multilayer Perceptron achieved the lowest average error rate in both machine learning models, granting it the title of the top performing technique. The results displayed below highlight the individual performance of each machine learning technique in the top performing model, the Financial Futures-Market model.

CPI Food Category	Linear Regression	Multilayer Perceptron	SMOreg	M5P
Meat	22.9328	*2.5615	2.9485	8.6286
Seafood	*2.4149	5.548	5.6408	4.6865
Dairy	13.8861	*0.8778	8.0817	6.9907
Bakery	11.4088	*1.001	13.5076	4.4224
Fruit	2.8711	2.0861	*1.2892	2.1337
Vegetables	2.0273	5.6032	3.9283	*2.0273
Other	14.2377	*2.3415	10.0612	3.3396
Restaurants	1.1124	*0.5451	3.4897	0.7221
Average MAPE	9.78231	*4.412	6.80826	4.417

Table 3: Comparison of MAPE results of each technique in the Financial Future-Markets model. (* Lowest MAPE per CPI Food Category).

The evaluation of the top performing techniques does leave some questions unanswered such as what caused the Holt-Winters model to outperform the Linear Regression technique. The Holt-Winters model posted an average error of 8 while the Linear Regression technique was only able to achieve 9.8 in the top performing model. Additionally, the Average error rate of the M5P Tree technique is incredibly close to that of the top performing Multilayer Perceptron even though the M5P tree technique was only able to secure a top performance in a single category, vegetables.

5.4 Discussion of Results

Table 2, the comparison of 2016 CPI predicted values and 2016 CPI actual values, provides the best overall analysis of this research thesis. As the table indicates, the Financial Futures Market produced the best average results when comparing the predicted value to the actual CPI. At a 2.8% error rate the Financial Futures-Market model has a predictive value 1.5 times better than the Food Price Report model and 4.6 times better than the Holt-Winters model.

Table 1, the comparison of the lowest achieved MAPE score per CPI food category and irrespective of the machine learning technique, further illustrates the success of the Financial Future Model. The results in Table 1 are strikingly similar to those of Table 2, in that the Financial Futures Model again produces an error rate which is 1.5 time better than the Food Price Report model and 5 times better than the Holt-Winters model. Interestingly, the food categories which are predominantly produced domestically – Meat, Seafood, Dairy and bakery – were not bested by the model with significantly more domestic data, the Food Price Report Model. This hints at lack of sufficient domestic data for forecasting domestic food. Nevertheless, the Financial Futures-Market model having convincingly bested the other models in terms of training error and predictive value was selected as the top performing model.

With the top performing model selected the task then switched to determining if the model had any historical significance. Figure 1 visualizes these results by illustrating the errors produced by the model with working with historical data. The results are not as optimistic as those achieved in tables 1 and 2. According to figure 1, only the Bakery, Fruit and

Restaurant categories achieved downward sloping trend lines which indicate a reduced error rate with time. The remaining 5 categories each produced an upward sloping trend line to indicate increased errors with time. However, the Bakery, Fruit and Restaurant categories were among the best performing categories of the top performing Financial Future-Market model so it is reasonable to believe that the other categories could achieve downwards sloping trend lines if their error rates were slightly improved upon.

Finally, table 3 highlights the contributions of the Multilayer Perceptron to the Financial Futures-Market model by comparing all 4 techniques applied to that model. The Multilayer Perceptron bested 5 of the 8 categories to become the top performing technique but unfortunately its results for Seafood, Fruits, and Vegetables were among the worst MAPE scores across all the techniques. These three categories were bested by the other techniques though only the M5P Tree technique produced a MAPE score comparable to that of the Multilayer Perceptron.

5.3 Research Limitations

The research methodology is constrained by the WEKA workbench and the techniques it makes available as part of its suite of offering. WEKA was originally selected as the main tool of analysis because it was familiar to the researcher but this of course meant that the techniques used were limited only to those available in the WEKA suite. It is reasonable to believe more popular tools such the R programming language or deep learning software could also provide interesting results.

With a focus on the Canada Consumer Price Index, the results and conclusion are limited to population specific impacts. It would not be fair to assume a similar rate of success for other population specific regions so any impact from this research will likely only affect Canadians.

Finally data limitations have been restricted to the number of available instances from reliable data sources. The priority for the feature selection process was to provide as many potential attributes as possible, which means there was less of a priority placed on how many instances each data source offered. A conscious decision was made to ensure the instances dated as far back as the Dot Com era and this coincidentally became the data limit.

CHAPTER 6: CONCLUSION

As stated in section 1.2 Research Objectives, this research thesis aimed to:

1. Determine the top performing model of the three models assessed (Holt-Winters, Food Price Report, and Financial Futures-based Markets).
2. Determine the top performing machine learning technique of the four assessed (Linear Regression, Multilayer Perceptron, SMOreg, and M5P Tree).
3. Evaluate the performance of the Multilayer Perceptron as the only technique to incorporate backpropagation.

The top performing model, as identified in section 5.1 Forecast Performance, was the Financial Futures-Market based model with an average MAPE of 1.6 and a 2.8% error between the actual 2016 CPI values and the predicted 2016 CPI values. This outcome was hypothesized because price forecasting is an inherent feature of the futures market and was likely to have provided some statistical advantage over the Food Price Report and the benchmark models.

The top performing machine learning technique as identified in section 5.2 Historical Performance was the Multilayer Perceptron. It produced a 4.412 average MAPE though this was closely followed by the M5P Tree with a 4.417 average MAPE. The Multilayer Perceptron separated itself from the others by having achieved the lowest MAPE in 5 of the 8 CPI food categories in the top performing model, however the backtesting analysis did not yield as promising results. Only 3 CPI food categories, Bakery, Fruit and Restaurants, achieved a downward trending error during the period tested. Given that only 15 periods were available for the experiment and only 7 for the backtesting process, the model may have suffered from a lack of sufficient data points. Extending the time series could be a possible option for improving the historical accuracy of the Financial Futures-Market based model.

Initially, the following question was posed, “What is the best way to predict food prices for the average Canadian consumer?” A definitive answer to this question was never the goal of this research thesis but instead two key techniques for predicting food prices have been identified. Both the Correlation-based Feature Selection process and the Multilayer Perceptron have achieved significant success. The CFS process reduced the original Food Price Report dataset by as much as 98% and the original Financial Future-Market dataset as much as 85%. The Multilayer Perceptron, in addition to besting 5 of 8 categories in the Financial Futures-Market model as noted above, also won 4 of 8 categories in the Food Price Report model for a combined top performance twice as good as the other techniques.

There is room for this research to be expounded upon in the future to create even more accurate forecasts. Consideration can be given to incorporating more data instances to the experiment. The Food Price Report model and Financial Futures-Market model employed 17 and 15 instances respectively and though the literature indicates that this is a sufficient numbers of data points, the accuracy of the forecasts could be improved by the increasing the size of the datasets. Some consideration could also be given to identifying a more innovative dataset of features such the Volatility Index (VIX) which is a popular measure of market confidence. Polls results related to public confidence in the elected officials could also prove to be fruitful additions to the research datasets. Additionally, the machine learning techniques themselves could be adjusted or even substituted to improve their forecast accuracy and predictive value. In particular the accuracy of the Multilayer Perceptron technique could be greatly improved upon in the Seafood, Fruit and Vegetable categories. Finally, owing to the successful performances of both the Food Price Report and Financial Futures-Market based models, it is likely that even better forecasts can be achieved by combining financial and conventional datasets to share their strengths and increase the information gain. However, these suggestions are outside of the scope of this research thesis but do leave room for further research in this area.

BIBLIOGRAPHY

Akram, Q. (2009). Commodity prices, interest rates and the dollar. *Energy Economics*, 838-851.

Austen, I. (2016, 01 20). *In Canada, the 8-Dollar Cauliflower Shows the Pain of Falling Oil Prices.* Retrieved from New York Times: http://www.nytimes.com/2016/01/21/business/dealbook/in-canada-5-cauliflowers-cost-more-than-a-barrel-of-oil.html?_r=1

Bandaraa, J., & Caib, Y. (2014). The impact of climate change on food crop productivity, food prices and food security in South Asia. *Economic Analysis and Policy*, 451-465.

Billi , R., & Kahn, G. (2008). *What Is the Optimal Rate of Inflation?* Kansas City: Federal Reserve Bank of Kansas City.

Blomberg, S., & Harris, E. (1995). The Commodity–Consumer Price Connection: Fact or Fable? *Federal Reserve Board of New York Economic Policy Review*, pp. 21 - 38.

Bordo, M. (2008). *An Historical Perspective on the Crisis on 2007-2008.* Cambridge, MA: National Bureau of Economic Research.

Box, G., & Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

Bowman, C., & Husain, A. (2004). *Forecasting Commodity Prices: Futures Versus Judgment.* International Monetary Fund.

Cao, L., & Tay, F. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 1506 - 1518.

Carter, C., & Mohapatra, S. (2008, May). How Reliable Are Hog Futures as Forecasts? *American Journal of Agricultural Economics*, pp. 367 - 378.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 1247 - 1250.

Charlebois, S., Harris, J., Tyedmers, P., Bailey, M., Keselj, V., Conrad, C., . . . Chamberlain, S. (2016). *Canada Food Price Report 2017.* Halifax: Dalhousie University.

Chiang, W., Urban, T., & Baldridge, G. (1996). A neural network approach to mutual fund net asset value forecasting. *Omega*, 205-215.

Cody, B. J., & Mills, L. O. (1991). The Role of Commodity Prices in Formulating Monetary Policy. *The Review of Economics and Statistics*, 358-365.

de Faria, ,. E., Albuquerque, ,. M., Gonzalez, J., Cavalcante, ,. J., & Albuquerque, M. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications*, 12506-12509.

Emmons, W., & Yeager, T. (2002, January). *The Futures Market as Forecasting Tool: An Imperfect Crystal Ball.* St. Louis: Federal Reserve Bank of St. Louis - The Regional Economist.

European Union. (2015). *World food consumption patterns – trends and drivers* . DG Agriculture and Rural Development.

Food and Agriculture Organization. (2002). *Diet, Nutrition and the Prevention of Chronic Diseases.* Rome: Agriculture and Consumer Protection Department.

Food and Agriculture Organization. (2017). *Food Price Index.* FAO.

Food Banks Canada. (2016). *HungerCount 2016.* Toronto: Food Banks Canada.

Guoqiang Zhanga, M. Y. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 16-32.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning.* Hamilton: University of Waikato.

Hall, M. A. (2000). *Correlation-based Features Selection for discrete and Numeric Class Machine Learning.* Hamilton: University of Waikato.

Haofei, Z., Guoping, X., Fangting, Y., & Han, Y. (2007). A neural network model based on the multi-stage optimization approach for short-term food price forecasting in China. *Expert Systems with Applications*, 347-356.

Hecht-Nielsen, R. (1988). *Theory of the Backpropagation Neural Network.* La Jolla: University of Caliiomia at San Diego.

Ho, S., & Xie, M. (1998). The use of ARIMA model for reliability, forecatsing and analysis. *23rd International Conference on Computers and Industrial Engineering* (pp. 213-216). Great Britain: Pergamon.

Ho, Y.-C., & Bryson, A. (1969). *Applied Optimal Control.* New York: Blaisdell.

Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: principles and practice.* OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 679–688.

Joutz, F. L. (1997). Forecasting CPI Food Prices: An Assessment. *Agricultural & Applied Economics Association*, 1681-1685.

Joutz, F., Trost, R., Hallahan, C., Clauson, A., & Denbaly, M. (2000). *Retail Food Price Forecasting at ERS: The Process, Methodology and Performance from 1984 to*

1997. Washington, DC: The Economic Research Service, U.S. Depertment of Agriculture.

Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010). Comparative study of sttribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, pp. 271-277.

Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 307–319.

Lam, K. F., Mui, H. W., & Yuen, H. K. (2001). A note on minimizing absolute percentage error in combined forecasts. *Computers & Operations Research*, 1141–1147.

Malhotra, A., & Maloo, M. (2017). *Understanding Food Inflation in India: A Machine Learning Approach.* Bombay: Indian Institute of Technology.

Morrison, R. (2016, 08 15). *Food Prices and Spending.* Retrieved from United States Department of Agriculture Economic Research Service: http://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/food-prices-and-spending/

Nowrouz Kohzadi, M. S. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, 169-181.

Öner, C. (2012, 03). Inflation: Prices on the Rise. *Finance and Development*.

Organization of Economic Co-Operation and Development. (2008). *Rising food prices: Cause and consequences.* OECD.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 533-536.

Smart, M. (2011). *The Impact of Sales Tax Reform on Ontario Consumers: A First Look at the Evidence.* Calgary, AB: University of Calagary School of Public Policy.

Statistics Canada. (1996). *Your Guide to the Consumer Price Index.* Ottawa: Statistics Canada.

Statistics Canada. (2015). *Table 326-0031 - Basket Weights of the Consumer Price Index.* Statistics Canada.

Statistics Canada. (2016, 03 18). *Canadian Agri-Food and Seafood Exports by Country (by Value).* Retrieved from Agriculture and Agri-Food Canada: http://www.agr.gc.ca/eng/industry-markets-and-trade/statistics-and-market-information/agriculture-and-food-market-information-by-region/canadian-agri-food-and-seafood-exports-by-country-by-value/?id=1410072148226

Statistics Canada. (2016, 03 18). *Canadian Agri-Food and Seafood Imports by Country (by Value).* Retrieved from Agriculture and Agri-Food Canada: http://www.agr.gc.ca/eng/industry-markets-and-trade/statistics-and-market-information/agriculture-and-food-market-information-by-region/canadian-agri-food-and-seafood-imports-by-country-by-value/?id=1410072148225

Statistics Canada. (2016a). *Table 326-0009 - Average retail prices for gasoline and fuel oil.* Statistics Canada.

Statistics Canada. (2016b). *Table 326-0021 - Consumer Price Index.* Statistics Canada.

Statistics Canada. (2016c). *Table 282-0073 - Labour force survey estimates (LFS).* Statistics Canada.

T.H. Hann, E. S. (1996). Much ado about nothing? Exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data. *Neurocomputing*, 323-339.

Teranet. (2016, 12). *National Bank National Composite House Price Index.* Retrieved from House Price Index: http://www.housepriceindex.ca/default.aspx

The World Bank. (2015). *World Development Indicators: Structure of merchandise imports .* The World Bank.

Ticlavilca, A., Feuz,, D., & McKee, M. (2010). Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression. *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management.* St. Louis, MO .

Tomek, W. (1997). Commodity Futures Prices as Forecasts. *Review of Agricultural Economics*, pp. 23-44.

Tratar, L., & Strmčnik, E. (2016). The comparison of Holt–Winters method and Multiple regression method: A case study. *Energy*, 266-276.

Webb, R. (1988). *Commodity Prices As Predictors of Aggregate Price Change.* Richmond: Federal Reserve Bank of Richmond.

Wheeler, T., & Braun, J. (2013). Climate Change Impacts on Global Food Security. *Science*, 508-513.

Wilson, R. (2015). Food Price Report. *Research (University of Guelph)*, 17.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques.* Burlington, MA: Elsevier Inc.

Working, H. (1976). Futures Trading and Hedging. In *The Economics of Futures Trading* (pp. 68-82). Palgrave Macmillan UK.

Yu, L., & Liu, H. (2003). Feature Selection for High Dimensional Data: A Fast Correlation-based Filter Solution. *Proceedings, Twentieth International*

Conference on Machine Learning (pp. 856-863). Washington DC: Proceedings, Twentieth International Conference on Machine Learning.

Zhang, G., Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, 35-62.

APPENDIX A: DATA SOURCES – FOOD PRICE REPORT MODEL

Input	Duration	Source	Data Location
CAD vs USD	1999 - 2015	Canada Forex	Historical Interest rates
Canada Consumer Price Index Energy	1999 - 2015	Statistics Canada	Table 326-0020
Canada Deposable Income	1999 -2015	Statistics Canada	Table 380-0072
Canada Household Farm Income	1999 -2015	Statistics Canada	Table 380-0072
Canada Household Income	1999 -2015	Statistics Canada	Table 380-0072
Canada Household Net Saving	1999 -2015	Statistics Canada	Table 380-0072
Canada Immigrant Mean Income	1999 - 2014	Statistics Canada	Table 054-0018
Canada Immigrants with Income	1999 - 2014	Statistics Canada	Table 054-0018
Canada Income Distribution	2000 - 2014	Statistics Canada	Table 111-0008
Canada Net Lending or Borrowing	1999 - 2015	Statistics Canada	Table 380-0072
Unemployment Rate	2000 - 2015	Statistics Canada	Table 282-0086
Precipitation	1999 - 2015	Environmental Protection Agency	Climate Change Indicators
Temperature	1999 - 2015	Environmental Protection Agency	Climate Change Indicators
FAO Commodity Index	1999 - 2015	Food and Agriculture Organization	Food Price Index
Gross Agricultural Production	1999 - 2013	Food and Agriculture Organization	Value of Agriculture production
International Aid Dollars	1999 - 2015	Food and Agriculture Organization	Food Aid Shipment
Population	1999 - 2015	Food and Agriculture Organization	Annual Population
Agriculture Fertilizers	1999 - 2014	Food and Agriculture Organization	Fertilizers Trade Value
Canada Credit to Agriculture	1999 - 2015	Food and Agriculture Organization	Credit to Agriculture
US Credit to Agriculture	1999 - 2015	Food and Agriculture Organization	Credit to Agriculture
Crude Oil Prices	1999 - 2015	US Energy Information Administration	Crude Oil Prices: West Texas Intermediate
US Overnight Lending Rates	2000 - 2015	US Federal Reserve	

Input	Duration	Source	Data Location
Commodity Market Food Price	1999 - 2015	World Bank	Commodity Markets
Diesel Prices	1999 - 2014	World Bank	World Bank Development Indicators
Gasoline Prices	1999 - 2014	World Bank	World Bank Development Indicators

APPENDIX B: DATA SOURCES – FINANCIAL FUTURES-MARKET MODEL

Input	Duration	Source	Data Location
Soybeans(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Soybean Meal(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Soybean Oil(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Canola(WCE)	2001 - 2015	Moore Research Center	Historical Futures Prices
Corn(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Wheat(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Wheat(KCBT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Wheat(MGE)	2001 - 2015	Moore Research Center	Historical Futures Prices
Oats(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Rough Rice(CBOT)	2001 - 2015	Moore Research Center	Historical Futures Prices
Live Cattle(CME)	2001 - 2015	Moore Research Center	Historical Futures Prices
Feeder Cattle(CME)	2001 - 2015	Moore Research Center	Historical Futures Prices
Lean Hogs(CME)	2001 - 2015	Moore Research Center	Historical Futures Prices
Class III Milk(CME)	2001 - 2015	Moore Research Center	Historical Futures Prices
Cocoa(ICE)	2001 - 2015	Moore Research Center	Historical Futures Prices
Coffee "C"(ICE)	2001 - 2015	Moore Research Center	Historical Futures Prices
Orange Juice(ICE)	2001 - 2015	Moore Research Center	Historical Futures Prices
Sugar #11(ICE)	2001 - 2015	Moore Research Center	Historical Futures Prices
London Cocoa(LCE)	2001 - 2015	Moore Research Center	Historical Futures Prices
London Sugar(LCE)	2001 - 2015	Moore Research Center	Historical Futures Prices

APPENDIX C: STATIONARY TESTS

Stationarity Test: Meat

Dickey-Fuller test (ADF(stationary) / k: 2 / Meat):

Tau (Observed value)	1.854
Tau (Critical value)	-0.258
p-value (one-tailed)	0.999
alpha	0.05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 99.92%.

Phillips-Perron test (PP(intercept) / Lag: Short / Meat):

Tau (Observed value)	2.177
Tau (Critical value)	-0.258
p-value (one-tailed)	0.999
alpha	0.05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 99.92%.

KPSS test (Level / Lag Short / Meat):

Eta (Observed value)	1.555
Eta (Critical value)	0.463
p-value (one-tailed)	<0.0001
alpha	0.05

Test interpretation:
H0: The series is stationary.
Ha: The series is not stationary.
As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.
The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Seafood

Dickey-Fuller test (ADF(stationary) / k: 2 / Seafood):

Tau (Observed value)	-0.621
Tau (Critical value)	-0.258
p-value (one-tailed)	0.910
alpha	0.05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 90.98%.

Phillips-Perron test (PP(intercept) / Lag: Short / Seafood):

Tau (Observed value)	1.614
Tau (Critical value)	-0.258
p-value (one-tailed)	0.910
alpha	0.05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 90.98%.

KPSS test (Level / Lag Short / Seafood):

Eta (Observed value)	1.412
Eta (Critical value)	0.463
p-value (one-tailed)	<0.0001
alpha	0.05

Test interpretation:
H0: The series is stationary.
Ha: The series is not stationary.
As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.
The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Dairy

Dickey-Fuller test (ADF(stationary) / k: 2 / Dairy):

Tau (Observed value)	-1.049
Tau (Critical value)	-0.258
p-value (one-tailed)	0.833
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 83.30%.

Phillips-Perron test (PP(intercept) / Lag: Short / Dairy):

Tau (Observed value)	-1.351
Tau (Critical value)	-0.258
p-value (one-tailed)	0.833
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 83.30%.

KPSS test (Level / Lag Short / Dairy):

Eta (Observed value)	1.705
Eta (Critical value)	0.463
p-value (one-tailed)	< 0.0001
alpha	0.05

Test interpretation:

H0: The series is stationary.

Ha: The series is not stationary.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Bakery

Dickey-Fuller test (ADF(stationary) / k: 2 / Bakery):	
Tau (Observed value)	-1.776
Tau (Critical value)	-0.258
p-value (one-tailed)	0.596
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 59.62%.

Phillips-Perron test (PP(intercept) / Lag: Short / Bakery):	
Tau (Observed value)	-0.114
Tau (Critical value)	-0.258
p-value (one-tailed)	0.596
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 59.62%.

KPSS test (Level / Lag Short / Bakery):	
Eta (Observed value)	1.708
Eta (Critical value)	0.463
p-value (one-tailed)	< 0.0001
alpha	0.05

Test interpretation:

H0: The series is stationary.

Ha: The series is not stationary.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Fruit

Dickey-Fuller test (ADF(stationary) / k: 2 / Fruit):

Tau (Observed value)	-0.501
Tau (Critical value)	-0.258
p-value (one-tailed)	0.925
alpha	0.05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 92.51%.

Phillips-Perron test (PP(intercept) / Lag: Short / Fruit):

Tau (Observed value)	2.435
Tau (Critical value)	-0.258
p-value (one-tailed)	0.925
alpha	0.05

Test interpretation:
H0: There is a unit root for the series.
Ha: There is no unit root for the series. The series is stationary.
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 92.51%.

KPSS test (Level / Lag Short / Fruit):

Eta (Observed value)	1.606
Eta (Critical value)	0.463
p-value (one-tailed)	< 0.0001
alpha	0.05

Test interpretation:
H0: The series is stationary.
Ha: The series is not stationary.
As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.
The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Vegetables

Dickey-Fuller test (ADF(stationary) / k: 2 / Vegetables):

Tau (Observed value)	-1.244
Tau (Critical value)	-0.258
p-value (one-tailed)	0.785
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 78.45%.

Phillips-Perron test (PP(intercept) / Lag: Short / Vegetables):

Tau (Observed value)	0.834
Tau (Critical value)	-0.258
p-value (one-tailed)	0.785
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 78.45%.

KPSS test (Level / Lag Short / Vegetables):

Eta (Observed value)	1.524
Eta (Critical value)	0.463
p-value (one-tailed)	< 0.0001
alpha	0.05

Test interpretation:

H0: The series is stationary.

Ha: The series is not stationary.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Other

Dickey-Fuller test (ADF(stationary) / k: 2 / Other):

Tau (Observed value)	-2.459
Tau (Critical value)	-0.258
p-value (one-tailed)	0.296
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 29.57%.

Phillips-Perron test (PP(intercept) / Lag: Short / Other):

Tau (Observed value)	0.124
Tau (Critical value)	-0.258
p-value (one-tailed)	0.296
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 29.57%.

KPSS test (Level / Lag Short / Other):

Eta (Observed value)	1.694
Eta (Critical value)	0.463
p-value (one-tailed)	< 0.0001
alpha	0.05

Test interpretation:

H0: The series is stationary.

Ha: The series is not stationary.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Stationarity Test: Restaurants

Dickey-Fuller test (ADF(stationary) / k: 2 / Restaurants):

Tau (Observed value)	-1.961
Tau (Critical value)	-0.258
p-value (one-tailed)	0.513
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 51.28%.

Phillips-Perron test (PP(intercept) / Lag: Short / Restaurants):

Tau (Observed value)	1.198
Tau (Critical value)	-0.258
p-value (one-tailed)	0.513
alpha	0.05

Test interpretation:

H0: There is a unit root for the series.

Ha: There is no unit root for the series. The series is stationary.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

The risk to reject the null hypothesis H0 while it is true is 51.28%.

KPSS test (Level / Lag Short / Restaurants):

Eta (Observed value)	1.712
Eta (Critical value)	0.463
p-value (one-tailed)	< 0.0001
alpha	0.05

Test interpretation:

H0: The series is stationary.

Ha: The series is not stationary.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

APPENDIX D: HOLT-WINTERS TRIPLE EXPONENTIAL SMOOTHING

Holt-Winters: Meat

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options: -F "CANADA Consumer Price Index Meat "-L 1-M8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R2-8
Instances: 17
Attributes: 2

CANADA Consumer Price Index Meat:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Meat (1-step ahead)

inst#	actual	predicted	error
9	111.1	100.3602	-10.7398
10	113.4	93.7859	-19.6141
11	118.4	109.1828	-9.2172
12	119.1	111.8221	-7.2779

=== Predictions for test data: CANADA Consumer Price Index Meat (1-step ahead) ===

inst#	actual	predicted	error
13	125.4	108.8983	-16.5017
14	132	117.4474	-14.5526
15	134.8	127.1353	-7.6647
16	145.6	125.1395	-20.4605
17	156.6	134.044	-22.556

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Meat Annual Average
1	86.7
2	90.8
3	98.6
4	100
5	101.8
6	106.2
7	108.2
8	107.9
9	111.1
10	113.4
11	118.4
12	119.1
13*	108.8983

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Meat Annual Average
1	125.4
2	132
3	134.8
4	145.6
5	156.6
6*	137.5118

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Meat Annual Average	
N	4
Mean absolute error	11.7122
Root relative squared error	412.0646
Direction accuracy	66.6667
Relative absolute error	451.3643
Mean absolute percentage error	10.2147
Root mean squared error	12.6291
Mean squared error	159.4947

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Meat Annual Average	
N	5
Mean absolute error	16.3471
Root relative squared error	203.5808
Direction accuracy	75
Relative absolute error	209.0827
Mean absolute percentage error	11.6652
Root mean squared error	17.1478
Mean squared error	294.0474

Total number of instances: 5

Holt-Winters: Seafood

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Seafood" -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1,3-8
Instances: 17
Attributes: 2
CANADA Consumer Price Index Seafood:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Seafood (1-step ahead)

inst#	actual	predicted	error
9	100.1	98.0276	-2.0724
10	101.2	92.9181	-8.2819
11	108.6	99.0629	-9.5371
12	108.7	103.9772	-4.7228

=== Predictions for test data: CANADA Consumer Price Index Seafood (1-step ahead) ===

inst#	actual	predicted	error
13	109	107.0791	-1.9209
14	111.8	107.8925	-3.9075
15	114.3	112.2945	-2.0055
16	123.4	108.8868	-14.5132
17	125.4	115.8691	-9.5309

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Seafood
1	95.3
2	96
3	99.5
4	100
5	99.7
6	99.1
7	99.1
8	98
9	100.1
10	101.2
11	108.6
12	108.7
13*	107.0791

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Seafood
1	109
2	111.8
3	114.3
4	123.4
5	125.4
6*	106.5723

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Seafood	
N	4
Mean absolute error	6.1536
Root relative squared error	180.236
Direction accuracy	66.6667
Relative absolute error	262.1141
Mean absolute percentage error	5.8452
Root mean squared error	6.8218
Mean squared error	46.5365

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Seafood	
N	5
Mean absolute error	6.3756
Root relative squared error	178.2979
Direction accuracy	75
Relative absolute error	182.6653
Mean absolute percentage error	5.2747
Root mean squared error	8.0555
Mean squared error	64.8905

Total number of instances: 5

Holt-Winters: Dairy

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Dairy " -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1-2,4-8
Instances: 17
Attributes: 2

CANADA Consumer Price Index Dairy:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Dairy (1-step ahead)

inst#	actual	predicted	error
9	119.8	108.8124	-10.9876
10	124.5	109.9099	-14.5901
11	128.9	115.116	-13.784
12	130.2	119.3757	-10.8243

=== Predictions for test data: CANADA Consumer Price Index Dairy (1-step ahead) ===

inst#	actual	predicted	error
13	134	123.8376	-10.1624
14	135.9	121.7295	-14.1705
15	135.8	126.0307	-9.7693
16	135.5	127.8713	-7.6287
17	136.8	127.5372	-9.2628

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Dairy
1	92.4
2	93.9
3	97.1
4	100
5	103.1
6	105.7
7	111
8	115.5
9	119.8
10	124.5
11	128.9
12	130.2
13*	123.8376

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Dairy
1	134
2	135.9
3	135.8
4	135.5
5	136.8
6*	128.621

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Dairy	
N	4
Mean absolute error	12.5465
Root relative squared error	347.1974
Direction accuracy	100
Relative absolute error	376.9077
Mean absolute percentage error	9.9744
Root mean squared error	12.6566
Mean squared error	160.1905

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Dairy	
N	5
Mean absolute error	10.1988
Root relative squared error	902.9169
Direction accuracy	0
Relative absolute error	1134.2042
Mean absolute percentage error	7.5212
Root mean squared error	10.4261
Mean squared error	108.7028

Total number of instances: 5

Holt-Winters: Bakery

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Bakery e" -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1-3,5-8
Instances: 17
Attributes: 2
CANADA Consumer Price Index Bakery:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Bakery (1-step ahead)

inst#	actual	predicted	error
9	118.1	111.2656	-6.8344
10	132.4	109.2862	-23.1138
11	137.9	117.8414	-20.0586
12	138.8	122.0728	-16.7272

=== Predictions for test data: CANADA Consumer Price Index Bakery (1-step ahead) ===

inst#	actual	predicted	error
13	146	130.5614	-15.4386
14	150.4	136.5979	-13.8021
15	152.2	140.2117	-11.9883
16	151.4	142.8529	-8.5471
17	154.6	120.9061	-33.6939

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Bakery
1	94.7
2	94.4
3	97.3
4	100
5	104.3
6	107
7	109.8
8	113.6
9	118.1
10	132.4
11	137.9
12	138.8
13*	130.5614

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Bakery
1	146
2	150.4
3	152.2
4	151.4
5	154.6
6*	143.2765

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Bakery	
N	4
Mean absolute error	16.6835
Root relative squared error	227.2457
Direction accuracy	66.6667
Relative absolute error	289.3701
Mean absolute percentage error	12.4604
Root mean squared error	17.7701
Mean squared error	315.776

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Bakery	
N	5
Mean absolute error	16.694
Root relative squared error	678.7768
Direction accuracy	50
Relative absolute error	666.9748
Mean absolute percentage error	11.0135
Root mean squared error	18.8727
Mean squared error	356.1802

Total number of instances: 5

Holt-Winters: Fruit

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Fruit " -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1-4,6-8
Instances: 17
Attributes: 2
CANADA Consumer Price Index Fruit:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Fruit (1-step ahead)

inst#	actual	predicted	error
9	105.7	105.8745	0.1745
10	107.4	90.2157	-17.1843
11	113.1	99.489	-13.611
12	112	111.7322	-0.2678

=== Predictions for test data: CANADA Consumer Price Index Fruit (1-step ahead) ===

inst#	actual	predicted	error
13	115.9	108.8533	-7.0467
14	119.5	114.79	-4.71
15	122.3	112.4755	-9.8245
16	125.8	111.5164	-14.2836
17	132.4	120.0212	-12.3788

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Fruit
1	93.8
2	90.8
3	97.2
4	100
5	98.2
6	99.3
7	98.8
8	101.3
9	105.7
10	107.4
11	113.1
12	112
13*	108.8533

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Fruit
1	115.9
2	119.5
3	122.3
4	125.8
5	132.4
6*	116.4076

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Fruit	
N	4
Mean absolute error	7.8094
Root relative squared error	362.4308
Direction accuracy	33.3333
Relative absolute error	365.4491
Mean absolute percentage error	7.1098
Root mean squared error	10.962
Mean squared error	120.1656

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Fruit	
N	5
Mean absolute error	9.6487
Root relative squared error	249.2549
Direction accuracy	50
Relative absolute error	249.6786
Mean absolute percentage error	7.7517
Root mean squared error	10.253
Mean squared error	105.1237

Total number of instances: 5

Holt-Winters: Vegetables

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Vegetables" -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1-5,7-8
Instances: 17
Attributes: 2
CANADA Consumer Price Index Vegetables:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Vegetables (1-step ahead) ===

inst#	actual	predicted	error
9	98.6	94.6531	-3.9469
10	100.6	87.1711	-13.4289
11	110.2	84.9391	-25.2609
12	109.3	112.6835	3.3835

=== Predictions for test data: CANADA Consumer Price Index Vegetables (1-step ahead)

inst#	actual	predicted	error
13	117.1	110.8657	-6.2343
14	113.3	109.5159	-3.7841
15	117.4	104.3642	-13.0358
16	122.5	115.9342	-6.5658
17	130.8	116.2721	-14.5279

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Vegetables
1	86.9
2	88
3	93.3
4	100
5	94.3
6	92.1
7	93.6
8	98.4
9	98.6
10	100.6
11	110.2
12	109.3
13*	110.8657

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Vegetables
1	117.1
2	113.3
3	117.4
4	122.5
5	130.8
6*	104.6528

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Vegetables	
N	4
Mean absolute error	11.5051
Root relative squared error	292.5459
Direction accuracy	0
Relative absolute error	336.5873
Mean absolute percentage error	10.8426
Root mean squared error	14.5385
Mean squared error	211.3694

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Vegetables	
N	5
Mean absolute error	8.8296
Root relative squared error	186.424
Direction accuracy	75
Relative absolute error	177.9978
Mean absolute percentage error	7.2469
Root mean squared error	9.7702
Mean squared error	95.4574

Total number of instances: 5

Holt-Winters: Other

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Other " -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1-6,8
Instances: 17
Attributes: 2
CANADA Consumer Price Index Other:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Other (1-step ahead)

inst#	actual	predicted	error
9	110.1	106.5944	-3.5056
10	112.8	105.8441	-6.9559
11	120.5	109.3452	-11.1548
12	124.1	109.2337	-14.8663

=== Predictions for test data: CANADA Consumer Price Index Other (1-step ahead) ===

inst#	actual	predicted	error
13	128.2	117.7347	-10.4653
14	131	119.3055	-11.6945
15	129.9	124.5949	-5.3051
16	129.8	128.0782	-1.7218
17	133.7	126.3376	-7.3624

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Other
1	97
2	97.3
3	99
4	100
5	103.3
6	104.4
7	107
8	108.9
9	110.1
10	112.8
11	120.5
12	124.1
13*	117.7347

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Other
1	128.2
2	131
3	129.9
4	129.8
5	133.7
6*	117.4968

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Other	
N	4
Mean absolute error	9.1207
Root relative squared error	222.5137
Direction accuracy	33.3333
Relative absolute error	235.5501
Mean absolute percentage error	7.6468
Root mean squared error	10.0761
Mean squared error	101.5277

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Other	
N	5
Mean absolute error	7.3098
Root relative squared error	302.4929
Direction accuracy	25
Relative absolute error	330.1745
Mean absolute percentage error	5.6015
Root mean squared error	8.1437
Mean squared error	66.3194

Total number of instances: 5

Holt-Winters: Restaurants

Scheme: timeseries.HoltWinters -alpha 0.2 -beta 0.2 -gamma 0.2 cycle-length 1
Lagged and derived variable options:
-F "CANADA Consumer Price Index Restaurants " -L 1 -M 8
Relation: Holt-Winters-weka.filters.unsupervised.attribute.Remove-R1-7
Instances: 17
Attributes: 2
CANADA Consumer Price Index Restaurants:
Holt-Winters triple exponential smoothing.
Value smoothing factor: 0.2
Trend smoothing factor: 0.2
Seasonal smoothing factor: 0.2
Season cycle length: 1

=== Predictions for training data: CANADA Consumer Price Index Restaurants (1-step ahead) ===

inst#	actual	predicted	error
9	114.1	105.1372	-8.9628
10	117	106.3514	-10.6486
11	121.1	108.8561	-12.2439
12	124	113.4565	-10.5435

=== Predictions for test data: CANADA Consumer Price Index Restaurants (1-step ahead)

inst#	actual	predicted	error
13	127.5	116.3655	-11.1345
14	130.6	119.1692	-11.4308
15	132.6	122.6736	-9.9264
16	135.2	125.2877	-9.9123
17	138.9	128.1857	-10.7143

=== Future predictions from end of training data ===

inst#	CANADA Consumer Price Index Restaurants
1	92
2	94
3	96.9
4	100
5	102.5
6	105.2
7	108.2
8	111.1
9	114.1
10	117
11	121.1
12	124
13*	116.3655

=== Future predictions from end of test data ===

inst#	CANADA Consumer Price Index Restaurants
1	127.5
2	130.6
3	132.6
4	135.2
5	138.9
6*	128.7862

=== Evaluation on training data ===

Target	1-step-ahead
CANADA Consumer Price Index Restaurants	
N	4
Mean absolute error	10.5997
Root relative squared error	333.6922
Direction accuracy	100
Relative absolute error	337.7376
Mean absolute percentage error	8.8925
Root mean squared error	10.6631
Mean squared error	113.7008

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CANADA Consumer Price Index Restaurants	
N	5
Mean absolute error	10.6237
Root relative squared error	360.3378
Direction accuracy	100
Relative absolute error	368.279
Mean absolute percentage error	8.0033
Root mean squared error	10.6416
Mean squared error	113.2446

Total number of instances: 5

APPENDIX E: CORRELATION-BASED FEATURE SELECTION

Food Price Report Model: Meat

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,479-485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 4715
Merit of best subset found: 0.981

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Meat Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 35,43,92,93,147,162,165,166,173,179: 10
Commodity Price Aluminum real $/mt
Commodity Price Beef real $/kg
Commodity Price Meat chicken $/kg
Commodity Price Meat chicken real $/kg
Commodity Price Tobacco US import u.v. real $/mt
Global Aid Blended and Mix
Global Aid Coarse Grains
Global Aid Dried Fruits Total
Global Aid Other Non-Cereals
Population Africa

Food Price Report Model: Seafood

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478,480-485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 5714
Merit of best subset found: 0.983

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Seafood Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 35,43,49,92,93,141,146,241,324,337,467: 11
Commodity Price Aluminum real $/mt
Commodity Price Beef real $/kg
Commodity Price Cocoa real $/kg
Commodity Price Meat chicken $/kg
Commodity Price Meat chicken real $/kg
Commodity Price Tea avg 3 auctions real $/kg
Commodity Price Tobacco US import u.v. $/mt
Diesel Iran Islamic Rep
Diesel Yemen Rep
Gasoline Bolivia
Temperature

Food Price Report Model: Dairy

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478-479,481-485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 25536
Merit of best subset found: 0.996

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Dairy Products and Eggs Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes:
32,33,34,39,45,47,56,57,68,69,71,72,73,74,75,78,81,82,83,88,89,91,103,106,107,110,111,112,113,114,115,116,117,118,119,122,123,126,127,133,134,135,136,137,144,153,159,161,166,167,168,170,172,180,182,461: 56
Commodity Price Agriculture index 2010=100
Commodity Price Agriculture real index 2010=100
Commodity Price Aluminum $/mt
Commodity Price Barley real $/mt
Commodity Price Beverages real index 2010=100
Commodity Price Coal Australian real $/mt
Commodity Price Copper $/mt
Commodity Price Copper real $/mt
Commodity Price Fertilizers index 2010=100
Commodity Price Fertilizers real index 2010=100
Commodity Price Food real index 2010=100
Commodity Price Gold $/troy oz
Commodity Price Gold real $/troy oz
Commodity Price Grains index 2010=100
Commodity Price Grains real index 2010=100
Commodity Price Groundnuts $/mt
Commodity Price Iron ore cfr spot real $/dmtu
Commodity Price Lead $/mt

Commodity Price Lead real $/mt
Commodity Price Logs Malaysian $/cubic meter
Commodity Price Logs Malaysian real $/cubic meter
Commodity Price Maize real $/mt
Commodity Price Non-energy real index 2010=100
Commodity Price Other food index 2010=100
Commodity Price Other food real index 2010=100
Commodity Price Palm oil $/mt
Commodity Price Palm oil real $/mt
Commodity Price Phosphate rock $/mt
Commodity Price Phosphate rock real $/mt
Commodity Price Platinum $/troy oz
Commodity Price Platinum real $/troy oz
Commodity Price Potarsium chloride $/mt
Commodity Price Potarsium chloride real $/mt
Commodity Price Precious metals index 2010=100
Commodity Price Precious metals real index 2010=100
Commodity Price Rice Thai 5% $/mt
Commodity Price Rice Thai 5% real $/mt
Commodity Price Sawnwood Malaysian $/cubic meter
Commodity Price Sawnwood Malaysian real $/cubic meter
Commodity Price Soybean meal real $/mt
Commodity Price Soybean oil $/mt
Commodity Price Soybean oil real $/mt
Commodity Price Soybeans $/mt
Commodity Price Soybeans real $/mt
Commodity Price Tin $/mt
Commodity Price Wheat US HRW real $/mt
Cereals Price Index
Sugar Price Index
Global Aid Dried Fruits Total
Global Aid Edible Fat Total
Global Aid Fish & Products
Global Aid Milk Total
Global Aid Other Dairy Products
Population Americas
Population Europe
CANADA Consumer Price Index Energy

Food Price Report Model: Bakery

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478-480,482-485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 8873
Merit of best subset found: 0.999

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Bakery and Cereal Products Excluding Baby Food Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 28,48,73,77,99,106,107,116,117,122,123,132,133,182,183,476: 16
USA Unemployment 15 To 24 Years
Commodity Price Cocoa $/kg
Commodity Price Gold real $/troy oz
Commodity Price Groundnut oil real $/mt
Commodity Price Natural gas US real $/mmbtu
Commodity Price Other food index 2010=100
Commodity Price Other food real index 2010=100
Commodity Price Potarsium chloride $/mt
Commodity Price Potarsium chloride real $/mt
Commodity Price Rice Thai 5% $/mt
Commodity Price Rice Thai 5% real $/mt
Commodity Price Soybean meal $/mt
Commodity Price Soybean meal real $/mt
Population Europe
Population Oceania
Us Overnight Lending

Food Price Report Model: Fruit

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478-481,483-485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 5181
Merit of best subset found: 0.985

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Fruit, Fruit Preparations and Nuts Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 35,37,49,92,146,147,166,173,179,474: 10
Commodity Price Aluminum real $/mt
Commodity Price Banana US real $/kg
Commodity Price Cocoa real $/kg
Commodity Price Meat chicken $/kg
Commodity Price Tobacco US import u.v. $/mt
Commodity Price Tobacco US import u.v. real $/mt
Global Aid Dried Fruits Total
Global Aid Other Non-Cereals
Population Africa
Canada credit to agriculture

Food Price Report Model: Vegetables

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478-482,484-485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 6110
Merit of best subset found: 0.968

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Vegetables and Vegetable Preparations Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 4,35,37,49,93,146,147,166,173,179,180: 11
Household Net Saving
Commodity Price Aluminum real $/mt
Commodity Price Banana US real $/kg
Commodity Price Cocoa real $/kg
Commodity Price Meat chicken real $/kg
Commodity Price Tobacco US import u.v. $/mt
Commodity Price Tobacco US import u.v. real $/mt
Global Aid Dried Fruits Total
Global Aid Other Non-Cereals
Population Africa
Population Americas

Food Price Report Model: Other

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478-483,485
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 10695
Merit of best subset found: 0.996
Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Other Food Products and Nonalcoholic Beverages Annual Average):
CFS Subset Evaluator
Including locally predictive attributes
Selected attributes:
4,22,28,35,72,73,98,106,107,108,109,118,119,130,131,161,167,182,183,379,476: 21
Household Net Saving
Canada Unemployment 15 Years and Over
USA Unemployment 15 To 24 Years
Commodity Price Aluminum real $/mt
Commodity Price Gold $/troy oz
Commodity Price Gold real $/troy oz
Commodity Price Natural gas US $/mmbtu
Commodity Price Other food index 2010=100
Commodity Price Other food real index 2010=100
Commodity Price Other raw mat. index 2010=100
Commodity Price Other raw mat. real index 2010=100
Commodity Price Precious metals index 2010=100
Commodity Price Precious metals real index 2010=100
Commodity Price Silver $/troy oz
Commodity Price Silver real $/troy oz
Sugar Price Index
Global Aid Edible Fat Total
Population Europe
Population Oceania
Gasoline Iran Islamic Rep
Us Overnight Lending

Food Price Report Model: Restaurants

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Full Dataset-weka.filters.unsupervised.attribute.Remove-R1,478-484
Instances: 17
Attributes: 477
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 9330
Merit of best subset found: 1

Attribute Subset Evaluator (supervised, Class (numeric): 477 CANADA Consumer Price Index Food Purchased from Restaurants Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 35,72,73,76,78,79,107,116,117,144,146,147,172,179,181,182: 16
Commodity Price Aluminum real $/mt
Commodity Price Gold $/troy oz
Commodity Price Gold real $/troy oz
Commodity Price Groundnut oil $/mt
Commodity Price Groundnuts $/mt
Commodity Price Groundnuts real $/mt
Commodity Price Other food real index 2010=100
Commodity Price Potarsium chloride $/mt
Commodity Price Potarsium chloride real $/mt
Commodity Price Tin $/mt
Commodity Price Tobacco US import u.v. $/mt
Commodity Price Tobacco US import u.v. real $/mt
Global Aid Other Dairy Products
Population Africa
Population Asia
Population Europe

Financial futures-market based Model: Meat

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,23-29
Instances: 15
Attributes: 21
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 127
Merit of best subset found: 0.965

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Meat Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 12,18,19: 3
Feeder Cattle CME
Sugar 11 ICE
London Cocoa LCE

Financial futures-market based Model: Seafood

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22,24-29
Instances: 15
Attributes: 21
[list of attributes omitted]
Annual Average
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 109
Merit of best subset found: 0.926

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Seafood Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 12,18,19: 3
Feeder Cattle CME
Sugar 11 ICE
London Cocoa LCE

Financial futures-market based Model: Dairy

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22-23,25-29
Instances: 15
Attributes: 21
[list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 167
Merit of best subset found: 0.962

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Dairy Products and Eggs Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 2,10,11,12,14,15,19,20: 8
Soybean Meal CBOT
Rough Rice CBOT
Live Cattle CME
Feeder Cattle CME
Class III Milk CME
Cocoa ICE
London Cocoa LCE
London Sugar LCE

Financial futures-market based Model: Bakery

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22-24,26-29
Instances: 15
Attributes: 21
[list of attributes omitted]
Food Annual Average
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 150
Merit of best subset found: 0.969

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Bakery and Cereal Products Excluding Baby Food Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 2,10,12,18,19: 5
Soybean Meal CBOT
Rough Rice CBOT
Feeder Cattle CME
Sugar 11 ICE
London Cocoa LCE

Financial futures-market based Model: Fruit

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22-25,27-29
Instances: 15
Attributes: 21
[list of attributes omitted]
Average
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 151
Merit of best subset found: 0.944

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Fruit, Fruit Preparations and Nuts Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 10,11,12,18,19: 5
Rough Rice CBOT
Live Cattle CME
Feeder Cattle CME
Sugar 11 ICE
London Cocoa LCE

Financial futures-market based Model: Vegetables

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22-26,28-29
Instances: 15
Attributes: 21
[list of attributes omitted]
Annual Average
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 187
Merit of best subset found: 0.917

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Vegetables and Vegetable Preparations Annual Average):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 2,11,12,13,14,17,18,19: 8
Soybean Meal CBOT
Live Cattle CME
Feeder Cattle CME
Lean Hogs CME
Class III Milk CME
Orange Juice ICE
Sugar 11 ICE
London Cocoa LCE

Financial futures-market based Model: Other

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22-27,29
Instances: 15
Attributes: 21
              [list of attributes omitted]
Beverages Annual Average
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 139
        Merit of best subset found: 0.955

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Other Food Products and Nonalcoholic Beverages Annual Average):
        CFS Subset Evaluator
        Including locally predictive attributes

Selected attributes: 4,10,11,18,19: 5
                Canola WCE
                Rough Rice CBOT
                Live Cattle CME
                Sugar 11 ICE
                London Cocoa LCE

Financial futures-market based Model: Restaurants

Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,22-28
Instances: 15
Attributes: 21
    [list of attributes omitted]
Average
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
    Best first.
    Start set: no attributes
    Search direction: forward
    Stale search after 5 node expansions
    Total number of subsets evaluated: 141
    Merit of best subset found: 0.967

Attribute Subset Evaluator (supervised, Class (numeric): 21 CANADA Consumer Price Index Food Purchased from Restaurants Annual Average):
    CFS Subset Evaluator
    Including locally predictive attributes

Selected attributes: 10,11,18,19: 4
        Rough Rice CBOT
        Live Cattle CME
        Sugar 11 ICE
        London Cocoa LCE

APPENDIX F: FOOD PRICE REPORT MODEL

Top Performing Machine Learning Technique: Meat

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Lagged and derived variable options: -F "CPI Meat" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R1,43,51,100-101,155,170,173-174,181,187,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Meat|35|43|92|93|147|162|165|166|173|179)-V-weka.filters.unsupervised.attribute.Remove-R12
Instances: 17
Attributes: 11

=== Predictions for training data: CPI Meat (1-step ahead) ===

inst#	actual	predicted	conf	error
6	106.2	105.8373	102.933:106.2	-0.3627
7	108.2	108.3761	105.472:108.739	0.1761
8	107.9	109.3412	106.437:109.704	1.4412
9	111.1	111.568	108.664:111.931	0.468
10	113.4	115.4744	112.57:115.837	2.0744
11	118.4	118.1475	115.243:118.51	-0.2525
12	119.1	122.004	119.1:122.367	2.904
13	125.4	125.4052	122.501:125.768	0.0052
14	132	131.9897	129.086:132.352	-0.0103

=== Predictions for test data: CPI Meat (1-step ahead) ===

inst#	actual	predicted	conf	error
15	134.8	136.2917	133.388:136.654	1.4917
16	145.6	141.0255	138.121:141.388	-4.5745
17	156.6	147.8016	144.898:148.164	-8.7984

=== Future predictions from end of training data ===

inst#	CPI Meat
1	86.7
2	90.8
3	98.6
4	100
5	101.8
6	106.2
7	108.2
8	107.9
9	111.1
10	113.4
11	118.4

12	119.1
13	125.4
14	132
15*	136.2917

=== Future predictions from end of test data ===

inst#	CPI Meat
1	134.8
2	145.6
3	156.6
4*	151.1915

=== Evaluation on training data ===

Target	1-step-ahead
CPI Meat	
N	9
Mean absolute error	0.8549
Root relative squared error	34.3297
Direction accuracy	87.5
Relative absolute error	27.7717
Mean absolute percentage error	0.7504
Root mean squared error	1.3021
Mean squared error	1.6955

Total number of instances: 14

=== Evaluation on test data ===

Target	1-step-ahead
CPI Meat	
N	3
Mean absolute error	4.9549
Root relative squared error	64.3282
Direction accuracy	100
Relative absolute error	61.3437
Mean absolute percentage error	3.2889
Root mean squared error	5.7897
Mean squared error	33.5211

Total number of instances: 3

Top Performing Machine Learning Technique: Seafood

Scheme: SMOreg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"
Lagged and derived variable options: -F "CPI Seafood" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R2,43,51,57,100-101,149,154,249,332,345,475,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Seafood|35|43|49|92|93|141|146|241|324|337|467)-V-weka.filters.unsupervised.attribute.Remove-R13
Instances: 17
Attributes: 12
Number of kernel evaluations: 105 (98.122% cached)

=== Predictions for training data: CPI Seafood (1-step ahead) ===

inst#	actual	predicted	conf	error
6	99.1	99.1294	97.342:102.847	0.0294
7	99.1	99.0696	97.282:102.787	-0.0304
8	98	98.7854	96.998:102.503	0.7854
9	100.1	100.1284	98.341:103.846	0.0284
10	101.2	102.9871	101.2:106.704	1.7871
11	108.6	104.8826	103.096:108.6	-3.7174
12	108.7	108.6685	106.881:112.386	-0.0315
13	109	109.3355	107.548:113.053	0.3355
14	111.8	111.7791	109.992:115.496	-0.0209

=== Predictions for test data: CPI Seafood (1-step ahead) ===

inst#	actual	predicted	conf	error
15	114.3	112.2314	110.444:115.949	-2.0686
16	123.4	111.7266	109.94:115.444	-11.6734
17	125.4	116.0936	114.306:119.811	-9.3064

=== Future predictions from end of training data ===

inst#	CPI Seafood
1	95.3
2	96
3	99.5
4	100
5	99.7
6	99.1
7	99.1
8	98
9	100.1
10	101.2
11	108.6
12	108.7

```
13          109
14          111.8
15*         112.2314

=== Future predictions from end of test data ===
inst#   CPI Seafood
1          114.3
2          123.4
3          125.4
4*        117.0609

=== Evaluation on training data ===
Target                              1-step-ahead
================================================
CPI Seafood
  N                                 9
  Mean absolute error               0.7518
  Root relative squared error       50.5189
  Direction accuracy                87.5
  Relative absolute error           45.2123
  Mean absolute percentage error    0.715
  Root mean squared error           1.4042
  Mean squared error                1.9718

Total number of instances: 14

=== Evaluation on test data ===
Target                              1-step-ahead
================================================
CPI Seafood
  N                                 3
  Mean absolute error               7.6828
  Root relative squared error       160.2315
  Direction accuracy                50
  Relative absolute error           189.0072
  Mean absolute percentage error    6.2303
  Root mean squared error           8.7017
  Mean squared error                75.7187

Total number of instances: 3
```

Top Performing Machine Learning Technique: Dairy

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Lagged and derived variable options: -F "CPI Dairy" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R3,40-42,47,53,55,64-65,76-77,79-83,86,89-91,96-97,99,111,114-115,118-127,130-131,134-135,141-145,152,161,167,169,174-176,178,180,188,190,469,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Dairy|32|33|34|39|45|47|56|57|68|69|71|72|73|74|75|78|81|82|83|88|89|91|103|106|107|110|111|112|113|114|115|116|117|118|119|122|123|126|127|133|134|135|136|137|144|153|159|161|166|167|168|170|172|180|182|461)-V-weka.filters.unsupervised.attribute.Remove-R58
Instances: 17
Attributes: 57

=== Predictions for training data: CPI Dairy (1-step ahead) ===

inst#	actual	predicted	conf	error
6	105.7	105.6345	104.55:106.385	-0.0655
7	111	110.2499	109.165:111	-0.7501
8	115.5	115.2237	114.139:115.974	-0.2763
9	119.8	120.2396	119.155:120.99	0.4396
10	124.5	124.7551	123.67:125.505	0.2551
11	128.9	128.4837	127.399:129.234	-0.4163
12	130.2	131.2849	130.2:132.035	1.0849
13	134	133.8522	132.767:134.602	-0.1478
14	135.9	136.2522	135.167:137.002	0.3522

=== Predictions for test data: CPI Dairy (1-step ahead) ===

inst#	actual	predicted	conf	error
15	135.8	138.3137	137.229:139.064	2.5137
16	135.5	140.5638	139.479:141.314	5.0638
17	136.8	142.8733	141.788:143.623	6.0733

=== Future predictions from end of training data ===

inst#	CPI Dairy
1	92.4
2	93.9
3	97.1
4	100
5	103.1
6	105.7
7	111
8	115.5
9	119.8
10	124.5
11	128.9
12	130.2

13	134
14	135.9
15*	138.3137

=== Future predictions from end of test data ===

inst#	CPI Dairy
1	135.8
2	135.5
3	136.8
4*	144.851

=== Evaluation on training data ===

Target	1-step-ahead
CPI Dairy	
N	9
Mean absolute error	0.4209
Root relative squared error	13.6736
Direction accuracy	100
Relative absolute error	12.3258
Mean absolute percentage error	0.3416
Root mean squared error	0.5162
Mean squared error	0.2664

Total number of instances: 14

=== Evaluation on test data ===

Target	1-step-ahead
CPI Dairy	
N	3
Mean absolute error	4.5503
Root relative squared error	592.6853
Direction accuracy	50
Relative absolute error	696.0686
Mean absolute percentage error	3.3426
Root mean squared error	4.7905
Mean squared error	22.9485

Total number of instances: 3

Top Performing Machine Learning Technique: Bakery

Scheme: M5P -M 4.0
Lagged and derived variable options: -F "CPI Bakery" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R4,36,56,81,85,107,114-115,124-125,130-131,140-141,190-191,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Bakery|28|48|73|77|99|106|107|116|117|122|123|132|133|182|183|476)-V
Instances: 17
Attributes: 17

=== Predictions for training data: CPI Bakery (1-step ahead) ===

inst#	actual	predicted	conf	error
6	107	109.3588	105.519:112.815	2.3588
7	109.8	113.3996	109.559:116.856	3.5996
8	113.6	117.4403	113.6:120.897	3.8403
9	118.1	121.4811	117.641:124.938	3.3811
10	132.4	129.8662	126.026:133.323	-2.5338
11	137.9	134.4435	130.603:137.9	-3.4565
12	138.8	139.0209	135.181:142.477	0.2209
13	146	143.5983	139.758:147.055	-2.4017
14	150.4	148.1757	144.335:151.632	-2.2243

=== Predictions for test data: CPI Bakery (1-step ahead) ===

inst#	actual	predicted	conf	error
15	152.2	152.7531	148.913:156.21	0.5531
16	151.4	157.3305	153.49:160.787	5.9305
17	154.6	161.9079	158.068:165.364	7.3079

=== Future predictions from end of training data ===

inst#	CPI Bakery
1	94.7
2	94.4
3	97.3
4	100
5	104.3
6	107
7	109.8
8	113.6
9	118.1
10	132.4
11	137.9
12	138.8
13	146
14	150.4
15*	152.7531

=== Future predictions from end of test data ===

inst#	CPI Bakery
1	152.2
2	151.4
3	154.6
4*	166.4853

=== Evaluation on training data ===

Target	1-step-ahead
CPI Bakery	
N	9
Mean absolute error	2.6686
Root relative squared error	44.2102
Direction accuracy	100
Relative absolute error	49.9035
Mean absolute percentage error	2.1588
Root mean squared error	2.8642
Mean squared error	8.2036

Total number of instances: 14

```
=== Evaluation on test data ===
Target                              1-step-ahead
==================================================
CPI Bakery
  N                                 3
  Mean absolute error               4.5972
  Root relative squared error       285.3274
  Direction accuracy                50
  Relative absolute error           330.9595
  Mean absolute percentage error    3.0025
  Root mean squared error           5.4431
  Mean squared error                29.6273
Total number of instances: 3
```

Top Performing Machine Learning Technique: Fruit

Scheme: SMOreg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"
Lagged and derived variable options: -F "CPI Fruit" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R5,43,45,57,100,154-155,174,181,187,482,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Fruit|35|37|49|92|146|147|166|173|179|474)-V-weka.filters.unsupervised.attribute.Remove-R12
Instances: 17
Attributes: 11
Number of kernel evaluations: 105 (96.956% cached)

=== Predictions for training data: CPI Fruit (1-step ahead) ===

inst#	actual	predicted	conf	error
6	99.3	99.1635	98.588:102.068	-0.1365
7	98.8	98.8079	98.232:101.713	0.0079
8	101.3	101.8665	101.291:104.771	0.5665
9	105.7	105.6448	105.069:108.55	-0.0552
10	107.4	107.452	106.877:110.357	0.052
11	113.1	110.1951	109.62:113.1	-2.9049
12	112	112.5754	112:115.48	0.5754
13	115.9	114.7777	114.202:117.683	-1.1223
14	119.5	119.5552	118.98:122.46	0.0552

=== Predictions for test data: CPI Fruit (1-step ahead) ===

inst#	actual	predicted	conf	error
15	122.3	122.6087	122.033:125.514	0.3087
16	125.8	128.2565	127.681:131.161	2.4565
17	132.4	131.6707	131.095:134.576	-0.7293

=== Future predictions from end of training data ===

inst#	CPI Fruit
1	93.8
2	90.8
3	97.2
4	100
5	98.2
6	99.3
7	98.8
8	101.3
9	105.7
10	107.4
11	113.1
12	112

13 115.9
14 119.5
15* 122.6087

=== Future predictions from end of test data ===

inst#	CPI Fruit
1	122.3
2	125.8
3	132.4
4*	137.0111

=== Evaluation on training data ===

Target	1-step-ahead
CPI Fruit	
N	9
Mean absolute error	0.6084
Root relative squared error	33.8094
Direction accuracy	87.5
Relative absolute error	22.8175
Mean absolute percentage error	0.5447
Root mean squared error	1.0738
Mean squared error	1.153

Total number of instances: 14

=== Evaluation on test data ===

Target	1-step-ahead
CPI Fruit	
N	3
Mean absolute error	1.1648
Root relative squared error	34.3009
Direction accuracy	100
Relative absolute error	31.5426
Mean absolute percentage error	0.9187
Root mean squared error	1.4901
Mean squared error	2.2205

Total number of instances: 3

Top Performing Machine Learning Technique: Vegetables

Scheme: SMOreg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"
Lagged and derived variable options: -F "CPI Vegetables" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R6,12,43,45,57,101,154-155,174,181,187-188,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Vegetables|4|35|37|49|93|146|147|166|173|179|180)-V-weka.filters.unsupervised.attribute.Remove-R13
Instances: 17
Attributes: 12
Number of kernel evaluations: 105 (98.66% cached)

=== Predictions for training data: CPI Vegetables (1-step ahead) ===

inst#	actual	predicted	conf	error
6	92.1	92.0507	91.029:98.148	-0.0493
7	93.6	93.6552	92.634:99.752	0.0552
8	98.4	98.3415	97.32:104.438	-0.0585
9	98.6	98.642	97.62:104.739	0.042
10	100.6	101.6216	100.6:107.718	1.0216
11	110.2	104.4997	103.478:110.597	-5.7003
12	109.3	109.3585	108.337:115.455	0.0585
13	117.1	111.0031	109.982:117.1	-6.0969
14	113.3	113.3168	112.295:119.414	0.0168

=== Predictions for test data: CPI Vegetables (1-step ahead) ===

inst#	actual	predicted	conf	error
15	117.4	120.2023	119.181:126.299	2.8023
16	122.5	121.5262	120.505:127.623	-0.9738
17	130.8	130.9399	129.918:137.037	0.1399

=== Future predictions from end of training data ===

inst#	CPI Vegetables
1	86.9
2	88
3	93.3
4	100
5	94.3
6	92.1
7	93.6
8	98.4
9	98.6
10	100.6
11	110.2
12	109.3

13	117.1
14	113.3
15*	120.2023

=== Future predictions from end of test data ===

inst#	CPI Vegetables
1	117.4
2	122.5
3	130.8
4*	132.9053

=== Evaluation on training data ===

Target	1-step-ahead
CPI Vegetables	
N	9
Mean absolute error	1.4555
Root relative squared error	59.8279
Direction accuracy	75
Relative absolute error	42.6464
Mean absolute percentage error	1.2975
Root mean squared error	2.8032
Mean squared error	7.8582

Total number of instances: 14

=== Evaluation on test data ===

Target	1-step-ahead
CPI Vegetables	
N	3
Mean absolute error	1.3053
Root relative squared error	10.0991
Direction accuracy	100
Relative absolute error	8.3111
Mean absolute percentage error	1.0963
Root mean squared error	1.7147
Mean squared error	2.9404

Total number of instances: 3

Top Performing Machine Learning Technique: Other

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Lagged and derived variable options: -F "CPI Other" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R7,12,30,36,43,80-81,106,114-117,126-127,138-139,169,175,190-191,387,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Other|4|22|28|35|72|73|98|106|107|108|109|118|119|130|131|161|167|182|183|379|476)-V
Instances: 17
Attributes: 22
=== Predictions for training data: CPI Other (1-step ahead) ===

inst#	actual	predicted	conf	error
6	104.4	104.4428	102.641:105.678	0.0428
7	107	106.8118	105.01:108.047	-0.1882
8	108.9	108.5265	106.725:109.762	-0.3735
9	110.1	111.9016	110.1:113.137	1.8016
10	112.8	114.2486	112.447:115.484	1.4486
11	120.5	119.265	117.463:120.5	-1.235
12	124.1	124.478	122.676:125.713	0.378
13	128.2	128.0076	126.206:129.243	-0.1924
14	131	130.2419	128.44:131.477	-0.7581

=== Predictions for test data: CPI Other (1-step ahead) ===

inst#	actual	predicted	conf	error
15	129.9	131.3906	129.589:132.626	1.4906
16	129.8	130.0565	128.255:131.291	0.2565
17	133.7	130.8503	129.049:132.085	-2.8497

=== Future predictions from end of training data ===

inst#	CPI Other
1	97
2	97.3
3	99
4	100
5	103.3
6	104.4
7	107
8	108.9
9	110.1
10	112.8
11	120.5
12	124.1
13	128.2
14	131
15*	131.3906

```
=== Future predictions from end of test data ===
inst#   CPI Other
1          129.9
2          129.8
3          133.7
4*       132.1246

=== Evaluation on training data ===
Target                                  1-step-ahead
================================================
CPI Other
  N                                     9
  Mean absolute error                   0.7131
  Root relative squared error           25.9294
  Direction accuracy                    100
  Relative absolute error                23.9675
  Mean absolute percentage error         0.6154
  Root mean squared error               0.931
  Mean squared error                    0.8668
Total number of instances: 14

=== Evaluation on test data ===
Target                                  1-step-ahead
================================================
CPI Other
  N                                     3
  Mean absolute error                   1.5323
  Root relative squared error           73.3411
  Direction accuracy                    100
  Relative absolute error               77.6552
  Mean absolute percentage error        1.1588
  Root mean squared error               1.8627
  Mean squared error                    3.4695
Total number of instances: 3
```

Top Performing Machine Learning Technique: Restaurants

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Lagged and derived variable options: -F "CPI Restaurants" -L 1 -M 5
Relation: Thesis Dataset NEWEST-weka.filters.unsupervised.attribute.Remove-V-R8,43,80-81,84,86-87,115,124-125,152,154-155,180,187,189-190,484-weka.filters.unsupervised.attribute.RemoveByName-E(CPI Restaurants|35|72|73|76|78|79|107|116|117|144|146|147|172|179|181|182)-V
Instances: 17
Attributes: 18

=== Predictions for training data: CPI Restaurants (1-step ahead) ===

inst#	actual	predicted	conf	error
6	105.2	105.0559	104.52:105.364	-0.1441
7	108.2	107.892	107.356:108.2	-0.308
8	111.1	110.9673	110.431:111.275	-0.1327
9	114.1	114.1905	113.654:114.498	0.0905
10	117	117.5363	117:117.844	0.5363
11	121.1	120.9157	120.379:121.224	-0.1843
12	124	124.2837	123.747:124.592	0.2837
13	127.5	127.4641	126.928:127.772	-0.0359
14	130.6	130.5322	129.996:130.84	-0.0678

=== Predictions for test data: CPI Restaurants (1-step ahead) ===

inst#	actual	predicted	conf	error
15	132.6	133.381	132.845:133.689	0.781
16	135.2	135.9838	135.447:136.292	0.7838
17	138.9	138.427	137.891:138.735	-0.473

=== Future predictions from end of training data ===

inst#	CPI Restaurants
1	92
2	94
3	96.9
4	100
5	102.5
6	105.2
7	108.2
8	111.1
9	114.1
10	117
11	121.1
12	124
13	127.5
14	130.6
15*	133.381

=== Future predictions from end of test data ===

inst#	CPI Restaurants
1	132.6
2	135.2
3	138.9
4*	140.5987

=== Evaluation on training data ===

Target	1-step-ahead
CPI Restaurants	
N	9
Mean absolute error	0.1981
Root relative squared error	8.0339
Direction accuracy	100
Relative absolute error	6.453
Mean absolute percentage error	0.1711
Root mean squared error	0.2471
Mean squared error	0.061

Total number of instances: 14

=== Evaluation on test data ===

Target	1-step-ahead
CPI Restaurants	
N	3
Mean absolute error	0.6792
Root relative squared error	20.2438
Direction accuracy	100
Relative absolute error	19.949
Mean absolute percentage error	0.5031
Root mean squared error	0.6947
Mean squared error	0.4827

Total number of instances: 3

APPENDIX G: FINANCIAL FUTURES-MARKET BASED MODEL

Top Performing Machine Learning Technique: Meat

=== Run information ===

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Lagged and derived variable options: -F "CPI Meat" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R2-20,22-26,29
Instances: 15
Attributes: 4

=== Predictions for training data: CPI Meat (1-step ahead) ===

inst#	actual	predicted	conf	error
6	107.9	108.185	107.475:109.285	0.285
7	111.1	110.8623	110.152:111.963	-0.2377
8	113.4	113.5769	112.867:114.677	0.1769
9	118.4	117.2996	116.589:118.4	-1.1004
10	119.1	119.8103	119.1:120.911	0.7103
11	125.4	126.0243	125.314:127.125	0.6243
12	132	131.2688	130.559:132.369	-0.7312

=== Predictions for test data: CPI Meat (1-step ahead) ===

inst#	actual	predicted	conf	error
13	134.8	138.1899	137.48:139.29	3.3899
14	145.6	144.8095	144.099:145.91	-0.7905
15	156.6	149.3546	148.644:150.455	-7.2454

=== Future predictions from end of training data ===

inst#	CPI Meat
1	98.6
2	100
3	101.8
4	106.2
5	108.2
6	107.9
7	111.1
8	113.4
9	118.4
10	119.1
11	125.4
12	132
13*	138.1899

=== Future predictions from end of test data ===

inst#	CPI Meat
1	134.8
2	145.6
3	156.6
4*	153.6055

=== Evaluation on training data ===

Target	1-step-ahead
CPI Meat	
N	7
Mean absolute error	0.5523
Root relative squared error	14.8151
Direction accuracy	100
Relative absolute error	14.8581
Mean absolute percentage error	0.4588
Root mean squared error	0.6334
Mean squared error	0.4013

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CPI Meat	
N	3
Mean absolute error	3.8086
Root relative squared error	47.2792
Direction accuracy	100
Relative absolute error	36.8617
Mean absolute percentage error	2.5615
Root mean squared error	4.6408
Mean squared error	21.5373

Total number of instances: 3

Top Performing Machine Learning Technique: Seafood

```
=== Run information ===

Scheme:
        LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4

Lagged and derived variable options:
        -F "CPI Seafood" -L 1 -M 5

Relation:     Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1,3-20,22-26,29
Instances:    15
Attributes:   4

CPI Seafood =
     0.8784 * ArtificialTimeIndex +
     0.0152 * ArtificialTimeIndex*Lag_CPI Seafood-5 +
    83.6036

=== Predictions for training data: CPI Seafood (1-step ahead) ===
     inst#     actual   predicted            conf       error
         6        98      97.949     96.394:101.482      -0.051
         7     100.1    100.3931     98.838:103.926      0.2931
         8     101.2    102.7551      101.2:106.288      1.5551
         9     108.6     105.067      103.512:108.6      -3.533
        10     108.7    107.4518    105.897:110.985     -1.2482
        11       109    109.6527    108.098:113.186      0.6527
        12     111.8    112.4039    110.849:115.937      0.6039

=== Predictions for test data: CPI Seafood (1-step ahead) ===
     inst#     actual   predicted            conf       error
        13     114.3    115.0213    113.466:118.554      0.7213
        14     123.4    119.0128    117.458:122.546     -4.3872
        15     125.4    121.5648     120.01:125.098     -3.8352

=== Future predictions from end of training data ===
inst#    CPI Seafood
1            99.5
2             100
3            99.7
4            99.1
5            99.1
6              98
7           100.1
8           101.2
9           108.6
```

10	108.7
11	109
12	111.8
13*	115.0213

=== Future predictions from end of test data ===

inst#	CPI Seafood
1	114.3
2	123.4
3	125.4
4*	124.1685

=== Evaluation on training data ===

Target	1-step-ahead
CPI Seafood	
N	7
Mean absolute error	1.1339
Root relative squared error	50.3721
Direction accuracy	100
Relative absolute error	57.1448
Mean absolute percentage error	1.0603
Root mean squared error	1.5738
Mean squared error	2.4768

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CPI Seafood	
N	3
Mean absolute error	2.9812
Root relative squared error	62.5424
Direction accuracy	100
Relative absolute error	74.0755
Mean absolute percentage error	2.4149
Root mean squared error	3.39
Mean squared error	11.4921

Total number of instances: 3

Top Performing Machine Learning Technique: Dairy

=== Run information ===

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Lagged and derived variable options: -F "CPI Dairy" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1-2,4-10,12-18,22,25-27
Instances: 15
Attributes: 9

=== Predictions for training data: CPI Dairy (1-step ahead) ===

inst#	actual	predicted	conf	error
6	115.5	115.8953	115.325:117.124	0.3953
7	119.8	119.6715	119.101:120.901	-0.1285
8	124.5	123.8539	123.283:125.083	-0.6461
9	128.9	127.6709	127.1:128.9	-1.2291
10	130.2	130.7707	130.2:132	0.5707
11	134	133.1961	132.625:134.425	-0.8039
12	135.9	134.8572	134.286:136.086	-1.0428

=== Predictions for test data: CPI Dairy (1-step ahead) ===

inst#	actual	predicted	conf	error
13	135.8	136.0167	135.446:137.246	0.2167
14	135.5	137.2755	136.705:138.505	1.7755
15	136.8	138.3918	137.821:139.621	1.5918

=== Future predictions from end of training data ===

inst#	CPI Dairy
1	97.1
2	100
3	103.1
4	105.7
5	111
6	115.5
7	119.8
8	124.5
9	128.9
10	130.2
11	134
12	135.9
13*	136.0167

=== Future predictions from end of test data ===

inst#	CPI Dairy
1	135.8
2	135.5
3	136.8
4*	139.3308

=== Evaluation on training data ===

Target	1-step-ahead
CPI Dairy	
N	7
Mean absolute error	0.6881
Root relative squared error	22.4175
Direction accuracy	100
Relative absolute error	21.6728
Mean absolute percentage error	0.5325
Root mean squared error	0.7709
Mean squared error	0.5944

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CPI Dairy	
N	3
Mean absolute error	1.1946
Root relative squared error	178.7297
Direction accuracy	50
Relative absolute error	210.4535
Mean absolute percentage error	0.8778
Root mean squared error	1.3824
Mean squared error	1.911

Total number of instances: 3

Top Performing Machine Learning Technique: Bakery

=== Run information ===

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Lagged and derived variable options: -F "CPI Bakery" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1-3,5-10,12-18,20,22-26,29
Instances: 15
Attributes: 6

=== Predictions for training data: CPI Bakery (1-step ahead) ===

inst#	actual	predicted	conf	error
6	113.6	107.9378	NaN:NaN	-5.6622
7	118.1	117.1326	NaN:NaN	-0.9674
8	132.4	127.2471	NaN:NaN	-5.1529
9	137.9	133.7129	NaN:NaN	-4.1871
10	138.8	137.3592	NaN:NaN	-1.4408
11	146	143.5934	NaN:NaN	-2.4066
12	150.4	145.8316	NaN:NaN	-4.5684

=== Predictions for test data: CPI Bakery (1-step ahead) ===

inst#	actual	predicted	conf	error
13	152.2	148.9076	NaN:NaN	-3.2924
14	151.4	152.3247	NaN:NaN	0.9247
15	154.6	154.2458	NaN:NaN	-0.3542

=== Future predictions from end of training data ===

inst#	CPI Bakery
1	97.3
2	100
3	104.3
4	107
5	109.8
6	113.6
7	118.1
8	132.4
9	137.9
10	138.8
11	146
12	150.4
13*	148.9076

=== Future predictions from end of test data ===

inst#	CPI Bakery
1	152.2
2	151.4
3	154.6
4*	156.2724

=== Evaluation on training data ===

Target	1-step-ahead
CPI Bakery	
N	7
Mean absolute error	3.4836
Root relative squared error	47.4932
Direction accuracy	100
Relative absolute error	50.8784
Mean absolute percentage error	2.6365
Root mean squared error	3.8881
Mean squared error	15.117

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CPI Bakery	
N	3
Mean absolute error	1.5238
Root relative squared error	30.021
Direction accuracy	50
Relative absolute error	31.9724
Mean absolute percentage error	1.001
Root mean squared error	1.985
Mean squared error	3.9401

Total number of instances: 3

Top Performing Machine Learning Technique: Fruit

=== Run information ===

Scheme: SMOreg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"

Lagged and derived variable options: -F "CPI Fruit" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1-4,6-18,22-26,29
Instances: 15
Attributes: 6

=== Predictions for training data: CPI Fruit (1-step ahead) ===

inst#	actual	predicted	conf	error
6	101.3	101.3426	101.3:103.403	0.0426
7	105.7	105.5578	105.515:107.618	-0.1422
8	107.4	107.3636	107.321:109.424	-0.0364
9	113.1	111.0394	110.997:113.1	-2.0606
10	112	112.0426	112:114.103	0.0426
11	115.9	115.902	115.859:117.963	0.002
12	119.5	119.4889	119.446:121.549	-0.0111

=== Predictions for test data: CPI Fruit (1-step ahead) ===

inst#	actual	predicted	conf	error
13	122.3	123.067	123.024:125.128	0.767
14	125.8	129.7883	129.746:131.849	3.9883
15	132.4	132.4929	132.45:134.553	0.0929

=== Future predictions from end of training data ===

inst#	CPI Fruit
1	97.2
2	100
3	98.2
4	99.3
5	98.8
6	101.3
7	105.7
8	107.4
9	113.1
10	112
11	115.9
12	119.5
13*	123.067

```
=== Future predictions from end of test data ===
inst#   CPI Fruit
1          122.3
2          125.8
3          132.4
4*       137.4613

=== Evaluation on training data ===
Target                                        1-step-ahead
================================================
CPI Fruit
  N                                           7
  Mean absolute error                         0.3339
  Root relative squared error                 22.5286
  Direction accuracy                          83.3333
  Relative absolute error                     11.2496
  Mean absolute percentage error              0.2974
  Root mean squared error                     0.7811
  Mean squared error                          0.6102
Total number of instances: 12

=== Evaluation on test data ===
Target                                        1-step-ahead
================================================
CPI Fruit
  N                                           3
  Mean absolute error                         1.6161
  Root relative squared error                 53.4015
  Direction accuracy                          100
  Relative absolute error                     40.4083
  Mean absolute percentage error              1.2892
  Root mean squared error                     2.3455
  Mean squared error                          5.5013
Total number of instances: 3
```

Top Performing Machine Learning Technique: Vegetables

=== Run information ===

Scheme: M5P -M 4.0

Lagged and derived variable options: -F "CPI Vegetables" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1-5,7-10,12-19,24-25,29
Instances: 15
Attributes: 9

=== Predictions for training data: CPI Vegetables (1-step ahead) ===

inst#	actual	predicted	conf	error
6	98.4	98.0326	94.858:101.854	-0.3674
7	98.6	100.6742	97.499:104.495	2.0742
8	100.6	103.7748	100.6:107.596	3.1748
9	110.2	106.379	103.204:110.2	-3.821
10	109.3	109.2703	106.096:113.091	-0.0297
11	117.1	113.5292	110.354:117.35	-3.5708
12	113.3	116.1209	112.946:119.942	2.8209

=== Predictions for test data: CPI Vegetables (1-step ahead) ===

inst#	actual	predicted	conf	error
13	117.4	120.45	117.275:124.271	3.05
14	122.5	122.4875	119.313:126.309	-0.0125
15	130.8	126.2562	123.081:130.077	-4.5438

=== Future predictions from end of training data ===

inst#	CPI Vegetables
1	93.3
2	100
3	94.3
4	92.1
5	93.6
6	98.4
7	98.6
8	100.6
9	110.2
10	109.3
11	117.1
12	113.3
13*	120.45

=== Future predictions from end of test data ===

inst#	CPI Vegetables
1	117.4
2	122.5
3	130.8
4*	130.4026

=== Evaluation on training data ===

Target	1-step-ahead
CPI Vegetables	
N	7
Mean absolute error	2.2655
Root relative squared error	53.7042
Direction accuracy	66.6667
Relative absolute error	63.7506
Mean absolute percentage error	2.0952
Root mean squared error	2.668
Mean squared error	7.118

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CPI Vegetables	
N	3
Mean absolute error	2.5354
Root relative squared error	46.6435
Direction accuracy	100
Relative absolute error	34.0027
Mean absolute percentage error	2.0273
Root mean squared error	3.1596
Mean squared error	9.983

Total number of instances: 3

Top Performing Machine Learning Technique: Other

=== Run information ===

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Lagged and derived variable options: -F "CPI Other" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1-6,8-12,14-18,21-26,29
Instances: 15
Attributes: 6

=== Predictions for training data: CPI Other (1-step ahead) ===

inst#	actual	predicted	conf	error
6	108.9	108.4894	108.453:110.316	-0.4106
7	110.1	109.9971	109.961:111.824	-0.1029
8	112.8	112.3456	112.309:114.172	-0.4544
9	120.5	118.6734	118.637:120.5	-1.8266
10	124.1	124.1361	124.1:125.963	0.0361
11	128.2	127.6522	127.616:129.479	-0.5478
12	131	129.5962	129.56:131.423	-1.4038

=== Predictions for test data: CPI Other (1-step ahead) ===

inst#	actual	predicted	conf	error
13	129.9	129.146	129.11:130.973	-0.754
14	129.8	127.2183	127.182:129.045	-2.5817
15	133.7	127.7435	127.707:129.57	-5.9565

=== Future predictions from end of training data ===

inst#	CPI Other
1	99
2	100
3	103.3
4	104.4
5	107
6	108.9
7	110.1
8	112.8
9	120.5
10	124.1
11	128.2
12	131
13*	129.146

=== Future predictions from end of test data ===

inst# CPI Other

1	129.9
2	129.8
3	133.7
4*	128.6774

=== Evaluation on training data ===

Target	1-step-ahead
CPI Other	
N	7
Mean absolute error	0.6832
Root relative squared error	23.4848
Direction accuracy	100
Relative absolute error	19.7816
Mean absolute percentage error	0.5596
Root mean squared error	0.9254
Mean squared error	0.8563

Total number of instances: 12

=== Evaluation on test data ===

Target	1-step-ahead
CPI Other	
N	3
Mean absolute error	3.0974
Root relative squared error	166.4058
Direction accuracy	100
Relative absolute error	213.4559
Mean absolute percentage error	2.3415
Root mean squared error	3.7733
Mean squared error	14.238

Total number of instances: 3

Top Performing Machine Learning Technique: Restaurants

=== Run information ===

Scheme: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Lagged and derived variable options: -F "CPI Restaurants" -L 1 -M 5

Relation: Futures Dataset-weka.filters.unsupervised.attribute.Remove-R1-7,9-18,21-26,29
Instances: 15
Attributes: 5

=== Predictions for training data: CPI Restaurants (1-step ahead) ===

inst#	actual	predicted	conf	error
6	111.1	111.4266	111.085:111.741	0.3266
7	114.1	114.0807	113.739:114.395	-0.0193
8	117	117.2609	116.92:117.575	0.2609
9	121.1	120.7859	120.445:121.1	-0.3141
10	124	124.3412	124:124.655	0.3412
11	127.5	127.6932	127.352:128.007	0.1932
12	130.6	130.6382	130.297:130.952	0.0382

=== Predictions for test data: CPI Restaurants (1-step ahead) ===

inst#	actual	predicted	conf	error
13	132.6	133.1811	132.84:133.495	0.5811
14	135.2	135.4417	135.1:135.756	0.2417
15	138.9	137.4855	137.144:137.8	-1.4145

=== Future predictions from end of training data ===

inst#	CPI Restaurants
1	96.9
2	100
3	102.5
4	105.2
5	108.2
6	111.1
7	114.1
8	117
9	121.1
10	124
11	127.5
12	130.6
13*	133.1811

```
=== Future predictions from end of test data ===
inst#   CPI Restaurants
1            132.6
2            135.2
3            138.9
4*          139.2548

=== Evaluation on training data ===
Target                              1-step-ahead
==============================================
CPI Restaurants
  N                                 7
  Mean absolute error               0.2134
  Root relative squared error       7.0698
  Direction accuracy                100
  Relative absolute error           5.9845
  Mean absolute percentage error    0.1785
  Root mean squared error           0.2476
  Mean squared error                0.0613
Total number of instances: 12

=== Evaluation on test data ===
Target                              1-step-ahead
==============================================
CPI Restaurants
  N                                 3
  Mean absolute error               0.7458
  Root relative squared error       31.7325
  Direction accuracy                100
  Relative absolute error           26.289
  Mean absolute percentage error    0.5451
  Root mean squared error           0.8938
  Mean squared error                0.7989
Total number of instances: 3
```