Forecasting The Canadian Unemployment Rate

Comparison between a seasonal ARIMA model and an ARDL model with the spot price of a barrel of oil as an exogenous variable.

Omar Abdikader

(3760718)

Major Paper presented to the Department of Economics of the University of Ottawa in partial fulfillment of the requirements of the M.A. Economics

Supervisor: Professor Kathleen Day

Department of Economics

University of Ottawa

Faculty of Social Science

September 4, 2019

Acknowledgement

I would like to thank my supervisor, Prof. Kathleen Day, for patiently providing me with guidance, encouragement and advice throughout my time as her student. I have been extremely fortunate to have her as a supervisor who cared about my work, and who responded to my questions and inquiries promptly.

I would also like to thank Prof. Stuart Glosser for taking the time to teach me about time series analysis. If it were not for him, I would not be able to put my models together and finish my thesis. His input, and advise based on his knowledge and experience has taught me more about the subject of time series analysis. I will take this new found knowledge with me wherever I end up working.

I would like to thank the Department of Economics for giving me the opportunity to enroll and successfully complete my Master of Arts in Economics. I would also like to thank Prof. Abel Brodeur for inspiring me to carry out econometric research and for making the topic fun.

Finally, I would like to thank my family for the support and encouragement that they have provided me with throughout this endeavor.

Abstract

This purpose of this paper is to forecast the Canadian unemployment rate (UER) using a seasonal ARIMA model, a seasonal random walk model, and an autoregressive distributed lag (ARDL) model containing the West Texas Intermediate (WTI) spot price of oil as a leading indicator variable. The seasonal ARIMA or SARIMA model will be developed based on the Box-Jenkins method, and the seasonal random walk model will be determined based on the level of integration required to make the unemployment rate series stationary. One-step ahead and 3-steps ahead forecasts are carried out. The performance of each model is assessed based on three forecasting evaluation criteria: the root mean squared error (RMSE), the mean average (MAE), and the mean absolute percentage error (MAPE). The statistical significance of differences in forecast accuracy is then evaluated using the Diebold-Mariano test and the Clark-West test to determine the overall best model at each forecast horizon. The results obtained show that the models that most accurately forecast the Canadian unemployment rate are the ARIMA$(||1,3,5||,0,0)(0,1,1)_{12}$ and the ARIMA$(||3,4,5||,1,0)(0,1,1)_{12}$. Overall, the results obtained in this study provide little evidence for the use of the oil price as a leading indicator to forecast the Canadian unemployment rate due to the poor performance of the ARDL models compared to the ARIMA models.

Keywords: UER, SARIMA, seaonal random walk, ARDL, RMSE, MAE, MAPE.

Contents

1 Introduction

Forecasting of macroeconomic variables is an integral component of the policy development strategy of both public sector and private sector entities. The models economists develop in order to forecast macroeconomic variables depend on historical data as well as other information pertinent to the variable being studied. However, forecasting macroeconomic variables is a challenging endeavor. According to Stock and Watson (2017) economic variables are inherently difficult to forecast, and due to this inherent difficulty in economic forecasting, there has been very little progress made in forecasting over the last twenty years. The unemployment rate is a macroeconomic variable that is particularly challenging to forecast. According to Montgomery et al. (1998), the unemployment rate is challenging to forecast due to its forceful propensity to fall "counter-cyclically" during periods of business and economic expansions, and rise during periods of general economic contractions and business downturns.

Furthermore, a structural change in the economy may alter the trajectory of the unemployment rate variable and make it difficult to forecast. In Germany, a low unemployment rate during the 1960s was followed by two energy price shocks that elevated the unemployment rate, and although the unemployment rate declined when the shocks subsided it remained higher than the previous unemployment rate (Funke, 1992). The unemployment rate, according to Będowska-Sójka (2015), can also be affected by shocks that are induced by "institutional factors" such as labour taxes and employment protection policies.

Despite these challenges, since the unemployment rate is one of the macroeconomic variables along with GDP that provides the most accurate information on the overall health of the economy, forecasting it is of great importance. A rise in the unemployment rate can result in a loss of income, a loss of productivity, and place a greater burden on the fiscal budget due in part to a rise in unemployment benefits allocated for unemployed individuals (Dritsakis and Klazoglou, 2018). In fact, according to Dritsakis and Klazoglou (2018) a persistently high unemployment rate can result in economic inequality that will result in "redistributive pressure," which will create economic distortions that could impede growth and cause more unemployment. For this reason, being able to forecast the unemployment rate accurately might allow both public and private sector entities to develop policies to counteract the harmful effects of unemployment, and develop strategies that might enable the individuals who will be most affected by an increase in unemployment to prepare.

For the most part, unemployment rate forecasting models have been based on either historical time series properties of the unemployment rate, as well as according to Barnichon et al. (2012) the "near term indicators" of the labour market, or the relationship between unemployment changes and output growth known as Okun's law. Unemployment forecasting models, according to Jalles (2017), have also been shown to underestimate the unemployment rate during recessions, overestimate the unemployment rate during recoveries, and on average "over-predict" the unemployment rate, with this over-prediction decreasing at the onset of a recession. It is these types of factors that necessitate the careful testing of the unemployment rate and any macroeconomic variable series used to produce forecasts.

Throughout the literature, several papers have explored various methods of forecasting the unemployment rate, from univariate models to multivariate models as well as linear and non-linear models (Khan Jaffur et al., 2017). In each of these studies, the forecaster tested various models and determined the one that most accurately forecasted the unemployment rate (Khan Jaffur et al., 2017). Furthermore, the models used to carry out forecasts of the unemployment rate could differ between countries. Variables that could be used as exogenous variables in the forecasting model of country A may not be appropriate for forecasting the unemployment rate in country B, due in part to structural differences in their respective economies. For example, since the oil industry is significant in Canada and creates a lot of high paying jobs, it would make more sense to consider the inclusion of oil prices in the model used to forecast the Canadian unemployment rate. However, since a country such as Japan doesn't have an oil industry, it would not make sense to include oil prices as a leading indicator in a model used to forecast the Japanese unemployment rate.

Even global shocks that impact the entire world may not have the same impact on the unemployment rate in all of the countries. For instance, Claveria (2019) notes that the 2008 global recession and the euro debt crisis had varying effects on the progression of the unemployment rate across different European countries. According to Claveria (2019), after the onset of the global shocks, the unemployment rate in some European countries went up and temporarily peaked until it subsided down to natural levels, whereas in other European countries the unemployment rate gradually increased and maintained a high level, and after the global shocks ended, some countries continue to suffer from persistently high unemployment rate. What this shows is that for modelling purposes, the incorporation of additional information into the main model depends on the characteristics of the unemployment rate in individual countries. It does not make any sense to assume that the unemployment rate series follows the same trend in all countries.

The objective of this paper is to compare various models for forecasting the Canadian unemployment rate. Since the time series for the Canadian unemployment rate exhibits clear seasonal patterns, seasonal differencing is used in order to convert it into a series that can be modeled accurately. This will result in a seasonal random walk model, which will serve as the starting point for building the appropriate seasonal ARIMA model for this paper. However, because the unemployment rate also exhibits a slight trend, another ARIMA model that is both seasonally and nonseasonallly differenced will be evaluated. Next, an autoregressive distriubted lag (ARDL) model will be developed that contains the spot price of the barrel of oil as a leading indicator variable. The oil price variable is the spot price of a barrel of oil based on the West Texas Intermediate (WTI) price benchmark. one-step ahead and 3-steps ahead forecasts of the seasonal random walk model, the seasonal ARIMA model, the ARIMA model utilizing the twice differenced unemployment rate, and the ARDL model are carried out. The model that can most accurately forecast the Canadian unemployment rate at each horizon is chosen to be the best model.

2 Literature Review

Several papers have utilized time series models to forecast the unemployment rate. The most extensively used model in the literature is the ARIMA model, otherwise known as the Box-Jenkins model (Dritsakis and Klazoglou, 2018). Another time series model that has been used to forecast the unemployment rate is the VAR model. Furthermore, previous papers have used a number of leading indicators to forecast the unemployment rate. Some researchers also utilize a variety of manipulation strategies in order to account for the various external factors that could impact the direction of the unemployment rate and the accuracy of the forecasts. In all of these papers, the models chosen depend on the time series used as well as the variables incorporated in the models.

2.1 Models for forecasting the unemployment rate.

The three basic univariate time series models used to forecast the unemployment rate are the autoregressive (AR) model, the moving average (MA) model, and the ARIMA or Box-Jenkins model (Mahmudah, 2017). An ARIMA model is symbolized as ARIMA(p,d,q), where p represents the number of AR terms, d denotes the order of nonseasonal integration,[1] and q denotes the number of MA terms. Papers that utilize the ARIMA or Box-Jenkins model to forecast the unemployment rate are Montgomery et al. (1998), Mahmudah (2017), Funke (1992), Edlund and Karlsson (1993) and Dritsakis and Klazoglou (2018). Before the ARIMA method is used, it is crucial to examine the stationarity of the data.

One reason a time series is non-stationary is if it has a trend. According to Stock and Watson (2012), a time series can either exhibit a stochastic trend or a deterministic trend. Furthermore, according to Stock and Watson (2012), a macroeconomic variable such as the unemployment rate is more appropriately modeled as possessing a stochastic rather than a deterministic trend. If the underlying series exhibits non-stationarity in the form of a stochastic trend, a processes such as first differencing is used to make it stationary (Mahmudah, 2017).

A time series can also be non-stationary if it exhibits clear seasonal patterns. According to Hylleberg et al. (1990), a series might posses either a deterministic seasonal process, which can be made stationary by including seasonal dummies, or an integrated seasonal process, which can be made stationary by taking a seaonal difference. For example, Edlund and Karlsson (1993) use

[1]The term ‘order of integration’ and the term ‘order of differencing’ are the same and are used interchangeably throughout the paper.

seasonal differencing to convert some of the leading indicator variables into stationary variables before conducting any multivariate forecasting of the unemployment rate, and Dritsakis and Klazoglou (2018) use seasonal differencing to turn the unemployment rate series into a stationary series before modelling the series using the Box-Jenkins method. Similarly, Funke (1992) uses nonseasonal and seasonal differencing to turn the German unemployment rate into a stationary series.

Furthermore, if seasonal patterns still linger after the series has been made statationary, additional seasonal AR and MA terms can be added to the model as demonstrated by Dritsakis and Klazoglou (2018) and Edlund and Karlsson (1993). A seasonal ARIMA model is sybolized as ARIMA$(p,d,q)(P,D,Q)_s$ where p, d and q are defined as before, P denotes the number of seasonal AR terms, D denotes the order of seaonal integration, and Q denotes the number of seasonal MA terms (Dritsakis and Klazoglou, 2018). Montgomery et al. (1998) account for the seaonality present in the U.S. unemployment rate series by adding a seasonal multiplicative ARMA (4,4) factor to an ARIMA (1,1,0) model, while Funke (1992) adds a seasonal MA(Q) term to the nonseasonally and seasonally differenced German unemployment rate to obtain an ARIMA$(2,1,0)(0,1,1)_{12}$ model.

A very common multivariate time series model that is used to forecast the unemployment rate is the VAR model (Khan Jaffur et al., 2017). This model is utilized, for example, by Funke (1992) and Edlund and Karlsson (1993) in forecasting the German unemployment rate and the Swedish unemployment rate respectively. Another multivariate approach used to forecast the unemployment rate involves forecasting the unemployment rate using an ARIMA model with additional explanatory variables, as demonstrated by Claveria (2019). The transfer function model built by Edlund and Karlsson (1993), which treats the Swedish unemployment rate as a function of the logarithm of Swedish industrial production, is also an example of a multivariate model that can be used to forecast the unemployment rate.

Throughout the literature the performance of univariate as well as mutlivarite models are assessed based on which models achieve the lowest root mean squared errors (RMSE), mean average error (MAE), and mean average percentage error (MAPE). Claveria (2019) and Khan Jaffur et al. (2017) also use the Diebold-Mariano (DM) statistic to compare two forecasts and test whether one is significantly better than the other.

2.2 Leading indicators used to forecast the unemployment rate.

Authors use several leading indicators throughout the literature to forecast the unemployment rate. Funke (1992) uses the index of total industrial production (IP), the index of the volume of new orders (NO), and survey data on the future tendency of production (FT) in building a VAR model, while Edlund and Karlsson (1993) use real labour cost, the Swedish industrial production index, the consumer price index and real GDP as leading indicators in forecasting the unemployment rate using a VAR model.

In addition to using macroeconomic variables as leading indicators, behavioral variables such as expectations are also used as leading indicators of the unemployment rate. Claveria (2019) uses two leading indicators to forecast the unemployment rate. The first is a variable that Claveria (2019) designates as "the degree of agreement in consumer employment expectations," and the second is a variable that encapsulates the "measure of disagreement based on the dispersion of expectations." The data for both leading indicators are generated by a survey that asks consumers how they expect the unemployment rate to change in the coming 12 months. The first variable measures the level that consumers agree the unemployment rate will look like over the next 12 months, and the second variable measures the level of disagreement.

Although not utilized in the literature, a variable that can be used as a leading indicator to forecast the unemployment rate is the spot price of a barrel of oil. According to Jung and Das (2018), the decline in oil prices that ensued in 2014 caused an economic downturn in Canada. Since a drop in oil prices has had an adverse affect on the Canadian economy, there is a strong possibility that low oil prices may have had an impact on the Canadian unemployment rate, which would suggest that the price of oil could be used as a leading indicator in a model forecasting the Canadian unemployment rate.

Jung and Das (2018) examine the relationship between the price of oil and the unemployment rate in Canada and the United States. They use Granger causality tests and ordinary least squares (OLS) to evaluate the asymmetric impact of the spot price of oil, measured by the West Texas Intermediate (WTI) bench price, on the U.S. and Canadian unemployment rates. Jung and Das (2018) divide the sample into a post-technological boom period which runs from April 1995 to October 2016, and a pre-technological boom period which runs from January 1976 to March 1995.

The results of Jung and Das (2018) reveal that in the post-technological era, the oil price has a Granger causal impact on the Canadian unemployment rate. Furthermore, a negative asymmetrical relationship exists for oil prices and the unemployment rate in Canada. Also, three provinces that seem to be particularly affected are Newfoundland and Labrador, Alberta and Saskatchewan. Finally, the overall conclusion drawn from this study is that falling oil prices have an adverse affect on the unemployment rate in Canada. Therefore, the spot oil price seems like a suitable leading indicator that can be used to forecast the Canadian unemployment rate.

2.3 Review of results obtained for different studies.

Comparisons of unemployment rate forecasting models have been carried out for a variety of countries. Funke (1992) explores the forecasting accuracy of two univariate models and one multivariate model in forecasting the German monthly unemployment rate from 1961 to 1989. The study explores two univariate ARIMA models which are the multiple-impact ARIMA model and the standard Box-Jenkins ARIMA model. The multiple-impact ARIMA model Funke (1992) uses combines the standard univariate Box- Jenkins approach with multiple intervention/outlier detection methods to account for structural change and/or outliers present in the unemployment series, which, Funke (1992) asserts, are caused by a couple of price shocks that occurred in Germany during the 1960s which resulted in an unemployment rate that did not return to its original level.

The VAR model Funke (1992) constructs uses the following monthly series variables reflecting current and expected future economic activity: the index of total industrial production (IP), index of the volume of new orders (NO), and the survey data on the future tendency of production (FT).

In the end, Funke's results reveal that none of the three models was necessarily the best available at all forecast horizons. The two univariate techniques result in lower errors than the VAR model for short forecast horizons while they were inferior to the multivariate VAR model for longer forecast horizons. Also, there is not a lot of improvement in forecasting accuracy between the two univariate techniques. Finally, Funke (1992) determines that the multiple-impact ARIMA model outperforms the univariate ARIMA model because it has a better fit and is able to predict better based on Theil's inequality coefficient.

Edlund and Karlsson (1993) also use an ARIMA model and a number of VAR models to forecast the unemployment rate, in addition to using a transfer function model. The study examines the quarterly Swedish unemployment rate from 1964 to 1990. The VAR models for this study contain the following leading indicators: the logarithm of the index of Swedish industrial production, the logarithm of real GDP, the logarithm of consumer price index, the logarithm of OECD industrial production, and the logarithm of real labour cost. Like Funke (1992), Edlund and Karlsson (1993) also use a Box-Jenkins approach to build an ARIMA model. Finally, Edlund and Karlsson (1993) identify an appropriate transfer function model using a pre-whitneing method. This process involves building the transfer function model using the logarithm of Swedish industrial production as an explanatory variable in an equation forecasting the Swedish unemployment rate and determining which lags of the Swedish industrial production most accurately predicts the unemployment rate based on what is called the cross-correlation function (CCF).

Using the cross-correlation function in determining which lags of the independent variable determine the values of the dependent variable, however, requires the residuals of both the explantory and dependent variables to behave as white noise. If the residuals of the underlying series do not exhibit white noise, then the residuals are transformed to exhibit white noise. Edlund and Karlsson (1993) use seasonal differencing for both the Swedish unemployment rate and the Swedish industrial production series in order to ensure that their residuals behave as white noise so that their CCF can be examined in order to determine the appropriate number of lags. In the end, Edlund and Karlsson (1993) identify a transfer function model with a lag of zero to three quarters.

The results of Edlund and Karlsson (1993) reveal that the transfer function and ARIMA models produce the lowest RMSE for all forecasting horizons. For the VAR model, the additional variables are significant in predicting the first turning point in the Swedish unemployment rate series, but not the second turning point.[2] However, the ARIMA model was not very successful in forecasting the turning points, which leads Edlund and Karlsson (1993) to conclude that the ARIMA model does not appropriately capture the business cycle in its forecasts. However, since the VAR model was not that successful in predicting the second turning point in the unemployment rate series, Edlund and Karlsson (1993) conclude that some variables that need to be included in order to capture all of the cyclical events are missing. One variable Edlund and Karlsson (1993) hypothesize could have enabled the VAR model to predict all of the turning points is the Swedish exchange rate.

Montgomery et al. (1998) forecast both the quarterly and monthly U.S. unemployment rates covering the period from 1948 to 1993 using various linear and nonlinear time series models with a rolling estimator and compare their performance. The paper measures the forecasting performance of the models during economic contractions and expansions. The paper makes use of both U.S. quarterly and monthly unemployment rate series to determine which series achieves the greatest accuracy in forecasting the unemployment rate. Montgomery et al. (1998) make use of an ARMA model, a univariate ARIMA model, a non-linear threshold autoregressive (TAR) model, and a bivariate VARMA model which includes jobless claims as a leading indicator. The paper also compares forecasts with consensus forecasts from the Survey of Professional Forecasters (SPF). According to Montgomery et al. (1998), the SPF forecasts are group median forecasts originally conducted by the National Bureau of Economic Research (NBER) and the American Statistical Association (ASA), but now conducted by the Federal Reserve Bank of Philadelphia since 1990 on a quarterly basis.

Montgomery et al. (1998) use the ratio of mean squared errors (MSE) to make all comparisons, and the results show that the TAR models produce better forecasts than linear models in forecasting unemployment rates during periods of contraction and rapid increase in unemployment rates,

[2]According to Edlund and Karlsson (1993) the Swedish unemployment rate exhibited "two sharp turning points" during 1979-1990. These turning points are difficult to forecast and present challenges for the models

but not elsewhere. Specifically, the TAR model results in up to a 28 percent reduction in MSE for longer term forecasts. The bivariate VARMA model with quarterly initial claims outperforms the univariate benchmark linear model during periods of rapidly increasing unemployment. The best short-term forecasts are generated by a monthly bivariate ARMA model. Finally, the forecasts of nonlinear models combined with the consensus forecasts from the Survey of Professional Forecasters yield results that show significant improvements in forecasting accuracy over other existing methods. The reason this occurs, according to Montgomery et al. (1998), is that the SPF forecasts contain information on other exogenous variables which are not included in the simple univariate as well as TAR models. Hence, combining the SPF forecasts with forecasts from other models yields more accurate forecasts.

Będowska-Sójka (2015) compares monthly unemployment rate forecasts from several models called Unobservable Component Models (UCM), which are basically multiple regression models with time varying coefficients. This study examines three baltic states: Estonia, Lithuania, and Latvia. The time period this study covers is from 2001 to 2014 and includes the 2008 financial crisis as well as both the steep increase and the gradual decrease in unemployment rates during the different phases of the business cycle. The results of Będowska-Sójka (2015) reveal that models that include the cyclical components have the best forecast accuracy. The structural model that contains the cyclical components produce superior forecasts of the unemployment rates; however, the advantage of these cyclical models diminishes for individual periods of unemployment decrease and increase.

Studies utilizing the Canadian unemployment rate are not common in the literature. However, a study which uses the seasonally adjusted Canadian unemployment rate covering the 1983-2013 period is conducted by Khan Jaffur et al. (2017). The paper examines the forecasting performance of non-linear and linear univariate time series models. Specifically, the models under examination are ARMA, AR, MA, GARCH, and EGARCH models (EGARCH stands for exponential generalized autoregressive conditional heteroscedasticity), and the Diebold-Mariano test is used to select the best model for each forecast horizon.

The results of Khan Jaffur et al. (2017) reveal that the ARMA(3,3) model is the best model for 3 to 9 months ahead forecasts, an ARMA(3,4) model is the best for 12 months ahead forecasts, a MA(1)-GARCH(1,1) model for 24 months ahead forecasts and a MA(5) model for the rest of the forecast horizons. Furthermore, according to the findings of the paper, the non-linear rather than the linear models are the ones that capture the asymmetry contained in the unemployment rate series at short and long forecast horizons.

Another paper that forecasts the U.S. unemployment rate using an ARIMA model built using the Box-Jenkins method is Dritsakis and Klazoglou (2018). The study forecasts the monthly U.S. unemployment rate covering a period which runs from January 1955 to July 2017. The paper evaluates the following models: the seasonal ARIMA model, which can also be expressed as a SARIMA model; the autoregressive conditional heteroskedasticity (ARCH) model; and the generalized autoregressive conditional heteroskedasticity (GARCH) model. The study evaluates the models individually as well as their combinations. Dritsakis and Klazoglou (2018) evaluate the performance of each model using the MAPE, the RMSE, and Theil's inequality coefficient. The results of Dritsakis and Klazoglou (2018) reveal that the model that best forecasts the U.S. unemployment rate is the SARIMA$(1,1,2)(1,1,1)_{12}$-GARCH(1,1) model.

A study by Claveria (2019) uses an ARIMAX model to forecast the unemployment rate. The primary benchmark model Claveria (2019) uses to conduct out of sample recursive forecasts of the unemployment rates of France, Austria, Greece, Italy, Netherlands, Portugal, German, and the UK is an ARIMA model. Claveria (2019) compares the forecasting performance of the ARIMA model with an ARIMAX model which includes "the degree of agreement in consumer employment expectations" as well as "a measure of disagreement based on the dispersion of expectations" as two leading indicators. Claveria's results reveal that both leading indicators in this paper enhance the accuracy of the unemployment rate forecasts of most of the countries investigated. Furthermore, the two leading indicator variables also lead to an an improvement in detecting turning points in the forecasts.

According to the results obtained throughout the literature, there is no clear consensus regarding which model is best at forecasting the unemployment rate. The success of the different models depends on what variables are used as well as the forecast horizon. Also, the type of leading indicators that are incorporated into the model as well as what type of multivariate model is required depend on not just the country the unemployment rate series is obtained from, but also the period being evaluated. However, what can be established from the literature is that the unemployment rate can exhibit a trend as well as a seasonal pattern, which does need to be taken into account during the model building phase. This current paper builds upon the foundations established in the literature by using an ARIMA model, developed through the Box-Jenkins approach, as well as a multivariate model, which in this paper is an ARDL model, to forecast the Canadian unemployment rate.

The seasonal ARIMA model that is used in this paper is the first seasonal univariate model for forecasting the Canadian unemployment rate. Furthermore, a paper forecasting the Canadian unemployment rate using an ARDL model with the spot price of the barrel of oil as an explanatory variable does not exist in the literature. Hence, this paper will be a first step in determining whether oil prices can be used to forecast the Canadian unemployment rate. The oil industry is a significant sector of the Canadian economy. If this paper establishes a relationship between the oil price and the unemployment rate in Canada, then perhaps the federal as well as the provincial governments will benefit from the findings contained in this paper to develop a policy supporting the oil industry in order to avert the harmful economic fallout of high unemployment not just in the oil sector but other sectors as well.

3 Data and Descriptive Statistics

The total sample range for the this study is 516 observations spanning January 1976 to December 2018. The main dependent variable being forecast in this study is the monthly Canadian unemployment rate obtained from Statistics Canada.[3] The graph in figure 1 shows the Canadian unem-

[3]The data is obtained from table 14-10-0287-01 (formerly CANSIM 282-0087). The data includes the seasonally unadjusted Canadian unemployment rate for both males and females aged 15 years and older. The link to the data is

ployment rate fluctuating and exhibiting a slight but not a very noticeable trend. Throughout the period being observed, the highest unemployment rate recorded is 14.1 percent and it occurred in March of 1983.

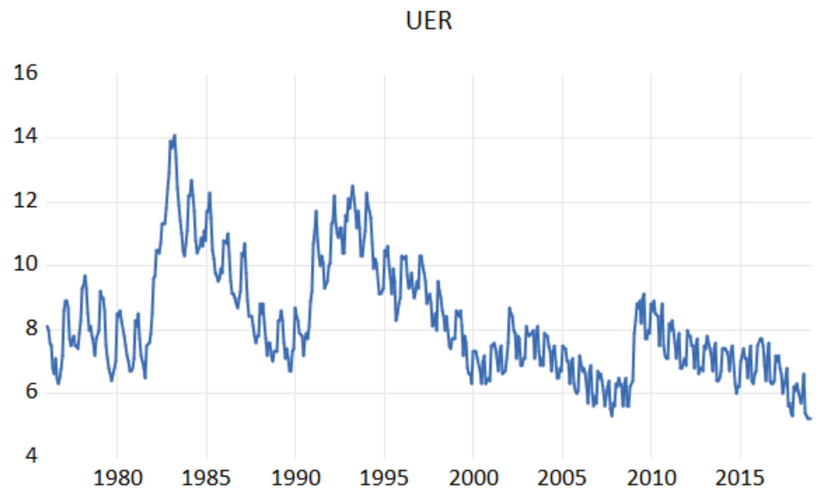


Figure 1: The Canadian unemployment rate from January 1976 to December 2018

The main leading indicator variable that is used in this paper is the spot price of oil, which is measured by the West Texas Intermediate (WTI) benchmark price. However, this is not the benchmark oil prices used in Canada. According to *Oil Sands Magazine*, the commonly used benchmarks in Canada are Canadian Light Sweet (CLS), priced out of Edmonton, Alberta, and Western Canadian Select (WCS), priced out of Hardisty, Alberta, with CLS being the price of light sweet crude oil, while WCS is the price of conventional heavy oil and diluted bitumen (Oil Sands Magazine, 2019).

According to *Oil Sands Magazine*, the WTI, which is priced out of Oklahoma's cushion storage hub, is priced higher than WCS and CLS because WTI is the price of light crude oil, which is oil that is easier to refine than heavy crude oil (Oil Sands Magazine, 2019). This price differential makes sense since heavy crude oil, which is more difficult to refine, has to be priced lower than light crude oil to make it more favorable for refiners to purchase. Hence, WTI is the most commonly

the following https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410028701 used oil price benchmark in North America, and both CLS and WCS are traded in reference to the WTI benchmark price (Oil Sands Magazine, 2019). Also since Jung and Das (2018) use the WTI spot price of oil in determining the casual impact of oil prices on Canadian unemployment, it makes sense to use this series as the main variable for oil prices in this paper.

Hence, the WTI spot price of oil will be used in this paper as the time series for the price of a barrel of oil, which is considered in this paper as a leading indicator of the Canadian unemployment rate. The monthly seasonally unadjusted data for the spot crude oil price per barrel of oil was obtained from the Federal Reserve Bank of St. Louis.[4]

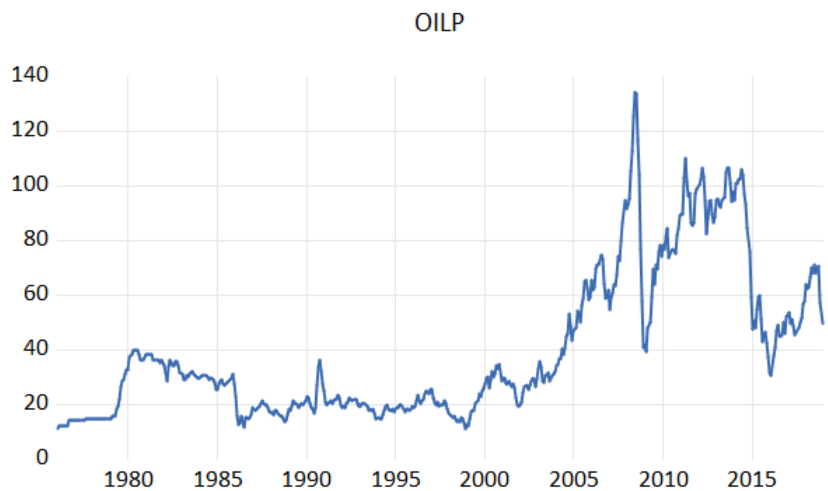


Figure 2: The spot crude oil price per barrel of oil from January 1976 to December 2018.

The data for this series, like that of the unemployment rate variable, covers the period from January 1976 to December 2018. The graph in figure 2 shows the oil price gradually increasing with a spike that occurs in June 2008. Furthermore, this series does not seem to exhibit the same fluctuations observed in the Canadian unemployment rate variable. In this paper, the observations used for the estimation of the parameters for all of the models utilized in this paper are January 1976 to December 2014. The observations set aside for the forecasting phase are January 2015 to

[4]The data are obtained from table WTISPLC of the economic research division. The variable is specifically called spot crude oil price: West Texas Intermediate (WTI), dollars per barrel. The weblink is the following https://fred.stlouisfed.org/series/WTISPLC.

December 2018.

Table 1: Descriptive Statistics.

	Variables				
	UER	OILP	SUER	TUER	DOILP
Mean	8.22	39.65	-0.0298	-0.001	0.0744
Median	7.80	29.33	-0.200	0.00	0.070
Maximum	14.10	133.93	4.800	1.30	13.36
Minimum	5.20	11.16	-1.900	-1.10	-27.25
Std. Dev	1.76	27.3	1.063	0.299	3.83
Observations	516	516	504	503	515

Table 1 contains the descriptive statistics for the variables used in this paper. UER represents the Canadian unemployment rate, OILP represents the WTI spot price of a barrel of oil, SUER represents the seasonally differenced unemployment rate, and DOILP represents the nonseasonally differenced price of the barrel of oil.[5] The time series for the unemployment rate variable, as shown in figure 1, contains a noticeable peak at around the beginning of the 1980s. This is expected since it coincides with a period in which the Canadian economy went through a deep recession. In fact, the recession of 1981-82, which affected both the United States and Canada, was the worst since the 1930s, and the cyclical downturn was more severe and sharper for Canada than it was for the United States (Wilson, 1985).

The graph for the unemployment rate shown in figure 1 contains other peaks that appear to coincide with business cycles of recession and economic growth; however, these recessions do not appear to be as severe as the recession in the early 1980s. Since that point in time, the unemployment rate series has been declining, and this could mean that this is a trend that could impact the assessment of the staionarity of the series. For this reason, it is prudent to use more powerful unit root tests in order to determine if the time series is non-stationary and the downward trend observed in figure 1 above is important.

[5] Since both variables used in this paper will be transformed in order to make them stationary, descriptive statistics for the transformed variables are necessary. The variable SUER $= UER_t - UER_{t-12}$, TUER $= SUER_t - SUER_{t-1}$, and DOILP $= OILP_t - OILP_{t-1}$.

4 Methodology

4.1 ARIMA Model Derivation

The first model that will be used to forecast the unemployment rate in this study is the ARIMA model. The ARIMA(p,d,q) model is basically a combination of an AR(p) model, and an MA(q) model, with the value for d being determined by the level of differencing that is required to convert the series into a stationary one. The parameters of the AR process can be estimated using the ordinary least squares method (OLS); however, if the series has MA terms or a combination of both MA and AR terms as in an ARMA process, then the maximum likelihood (ML) method is used to estimate the parameters (Dritsakis and Klazoglou, 2018).

In this paper the variable that will denote the unemployment rate is UER. A general AR process of order p that models the unemployment rate UER possesses the following form:

$$UER_t = \phi_0 + \phi_1 UER_{1-t} + \phi_2 UER_{2-t} + ... + \phi_p UER_{t-p} + \varepsilon_t \quad . \tag{1}$$

For the above equation, the white noise error is $\varepsilon_t \sim$ N(0,σ^2). A general MA model for the unemployment rate of order q where ε_t is the white noise error has the following expression:

$$UER_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q} \quad . \tag{2}$$

Next, an ARMA model, where ε_t is once again the white noise error, for the unemployment rate has the following form:

$$UER_t = \phi_0 + \phi_1 UER_{1-t} + \phi_2 UER_{2-t} + ... + \phi_p UER_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q} \quad . \tag{3}$$

Finally, an ARIMA model has the exact same form as the above ARMA model except it is constructed using a time series that has been differenced d times.[6] However, before the order of integration for a given time series is determined, the stationarity of the series must be assessed.

[6] d represents the order of integration.

A weakly stationary series is one in which the mean, variance and autocorrelation are constant over time. However, if these measures are not constant, then the series exhibits non-stationarity. Throughout the literature, the most common method of evaluating the stationarity of a series is to test for the presence of a unit root. A couple of tests that can be used to test for the presence of a unit root are the Augmented Dickey Fuller (ADF) test and the Philip-Perron (PP) test. For example, Dritsakis and Klazoglou (2018) use both of these tests to test for the presence of a unit root in the unemployment rate series for the Unite States.

However, although the ADF test and the PP test are commonly used throughout the literature, the conclusions drawn from these tests might be inaccurate. According to Ng and Perron (2001), many traditional unit root tests tend to over-reject the unit root hypothesis, because whenever the moving average polynomial of their series in first difference has a large negative root they experience severe size distortions, and whenever their autoregressive polynomial is close to but less than "unity" they have low power.[7] For example, according to Ng and Perron (2001), the Dickey Fuller GLS test, which is a unit root test that is conducted on a series that has been detrended by a generalized least square method, produces a t-statistic that has great power but it is not adequate enough to overcome size distortions that occur whenever the series being inspected contains an MA(1) component with a negative coefficient.

Furthermore, the lag length selection criteria used to conduct the unit root tests can also affect their power. For this reason, Ng and Perron (2001) proposed new modified lag length selection criteria, and investigated how these new criteria impacted the performance of a variety of unit root tests, such as the M tests developed by Perron and Ng (1996) which are called MZt, MZa, MSB, and MPT, and the DF-GLS test. The study conducted by Ng and Perron (2001) uses a class of Modified information criteria (MIC) containing "a penalty factor that is sample dependent" to improve the DF-GLS test. Furthermore, Ng and Perron (2001) use the modified Akaike information criterion (MAIC) along with data that has been detrended using GLS to produce MZt, MZa, MSB, and MPT tests that have good power and size.

[7]Statistical power is the probability of committing a type II error. If a test has low power, then the probability of obtaining a type II error is high and vice versa.

For this paper, six nonseasonal and one seasonal unit root test statistics are evaluated. The first unit root test that is conducted is the Augmented Dickey Fuller (ADF) test. Since this test is the most utilized unit root test in the unemployment rate forecasting literature, it is also carried out in this paper. However, since it is not as powerful as other unit root tests it will not be relied upon in this paper to be the sole determinant of the existence of a unit root. The results of the ADF test must be compared to the results of other unit root tests in order to determine whether the series is stationary or nonstationary. Although not as powerful as other tests, the ADF test is still reliable enough to shed some light on whether the series is stationary or nonstationary. It can serve as a starting point to evaluate the stationarity of the underlying series.

The second test that will be carried out is the DF-GLS test, followed by the four M tests called MZt, MZa, MSB, and MPT. For all the unit root tests, the information criterion that will be used is the MAIC criterion, because Ng and Perron (2001) show that the MAIC is superior to all other information criteria from both a numerical and theoretical standpoint. For each of these test statistics, their respective p-values must be lower than the critical value determined at either the 10 percent, 5 percent or the 1 percent level of significance in order to successfully reject the null hypothesis of the presence of a unit root.

The unit root tests discussed above only test for the presence of a unit root while igonoring seasonality. However, a time series can also be a nonstationary series if it exhibits clear seasonal patterns. To test for the presence of a seasonal unit root, this paper makes use of a seasonal unit root test called the HEGY test, developed by Hylleberg et al. (1990). The HEGY procedure tests for the presence of a seasonal unit root at specific seasonal frequencies depending on whether the underlying series is quarterly, monthly or annual. It is important to note however, that the HEGY test may not reveal a unit root at all of the frequencies under evaluation (Beaulieu and Miron, 1993). The underlying series is determined to be nonstationary if it contains a unit root at one or more seasonal frequencies.

In this paper, the seasonal ARIMA model that will be used as the one of the main univariate models in forecasting the unemployment rate is constructed using the Box-Jenkins methodology.

This approach consists of a model identification phase and a model diagnostic stage. The first step of the model identification stage starts with a careful assessment of the stationarity of the series. This process consists of two steps. The first step consists of a qualitative inspection of the graph of the raw data along with a careful examination of its correlogram. If the qualitative inspection of the graph of the raw series as well as its correlogram reveals a trend or a seasonal pattern, then a unit root test is conducted in order to verify whether or not the series in question is non-stationary.

After the stationarity of the series is assessed, and the series is determined to be nonstationary, the series is differenced in order to make it stationary, and the appropriate AR(p) and a MA(q) terms are determined from the assessment of the correlogram of the series in first differences. This step involves identifying any lagged values that have significant peaks outside the 95 percent confidence interval shown in the correlogram.[8] The appropriate lag values to be tested will then be chosen based on model parsimony. According to Ledolter and Abraham (1981), model parsimony, which consists of choosing a model with the fewest number of parameters, is a "desirable criterion" because such a models are simpler to explain and understand. Furthermore, according to Ledolter and Abraham (1981) each additional parameter placed in the model will result in more estimation variation.

After the appropriate number of lagged values are chosen based on the correlogram and model parsimony, various ARIMA models containing a combination of AR(p) and a MA(q) terms based on the chosen number of lags are estimated. The final choice of ARIMA model is based on which combination of AR(p) and MA(q) terms achieves the lowest AIC and SBIC. It is important to note that the sample period used for the estimation of the parameters in this paper is from January 1976 to December 2014. However, the sample that is used during the model selection process is adjusted depending on the maximum lag length chosen for the selection process. The final model adopted will use the entire sample set aside for the estimation of the parameters. It may have the same sample size as the sample size chosen for the model selection process depending on whether or not the chosen model is the one that contains the maximum lag length used during the selection

[8] Table 2.1 of chapter 2 of Enders (2014) contains the ACF and PACF for a number of ARMA processes providing guidelines for deriving the appropriate AR(p) and a MA(q) terms from a qualitative inspection of the correlogram.

process.

The ARIMA model derived at this stage is not the final model adopted. If the correlogram of the unemployment rate reveals that there are still significant peeks at higher lagged values, they will not be incorporated into the model during the model identification step due to model parsimony. Their inclusion in the model will be assessed during the model diagnostic stage.

The model diagnostic stage involves estimating the model that is selected during the model indentification step and examining the correlogram of the residuals to determine whether there is serial correlation. The purpose of this stage is to ensure the final model possesses residuals that are uncorrelated. The first part of this process involves a close examination of the correlogram of the residuals to determine if there are significant lagged values that need to be incorporated into the model. After those lagged values are determined, a new combination of AR(p) and MA(q) terms based on the chosen lagged values will be added to the ARIMA model derived during the model identification stage. Various ARIMA models containing the new combination of AR(p) and a MA(q) terms will be estimated and the final model will be chosen based on which combination of AR(p) and MA(q) terms achieves the lowest AIC and SBIC.

After the appropriate ARIMA model has been determined, the correlogram of the residuals is examined again to see if there are any more lagged values that need to be incorporated into the model. If the residuals contain no lagged values that are outside the 95 percent confidence interval, then the histogram of the residuals is assessed to see if they possess a normal distribution. After, the residuals of the new ARIMA model are assessed and found to contain no serial correlation, and are normally distributed, the model is adopted as the final ARIMA model and will be used as the bench mark model for this paper.

The next phase of this study involves the addition of all the appropriate variables including any seasonal AR(P) and a MA(Q) terms chosen for the final seasonal ARIMA model. The seasonal ARIMA model with the seasonally differenced unemployment rate denoted by SUER as the dependent variable along with all its seasonal terms possesses the following form:

$$SUER_t = \phi_0 + \sum_{i=1}^{p} \phi_i SUER_{t-i} + \sum_{i=1}^{P} \Phi_i SUER_{t-i\cdot s} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{j=1}^{Q} \Theta_j \varepsilon_{t-j\cdot s} + \varepsilon_t \quad . \tag{4}$$

The first two terms of equation 4 represent the appropriate number of AR(p) and a MA(q) terms used in this model with their respective coefficients being ϕ and θ. The next two terms represent any seasonal AR or seasonal MA terms that might need to be included in the model with their respective coefficients being Φ and Θ.[9]

4.2 ARDL Model Derivation

The second model that is utilized in this paper is an ARDL model. This model like the ARIMA model contains lagged values of the dependent variable as regressors. The primary difference between the ARDL model and the ARIMA model is that the ARDL also contains lagged values of an explanatory/exogenous variable as regressors. The derivation of the ARDL model like the ARIMA model involves two stages, the model identification stage and the model diagnostic stage. The modelling process begins with an assessment of the stationarity of the dependent and exogenous variables. Similar to the ARIMA model derivation process, this step of the ARDL process involves a qualitative inspection of the graph of the series along with its correlogram, followed by the appropriate unit root tests. After the stationarity of both series is ascertained, each series is then subjected to the appropriate level of differencing in order to make it stationary.

Before continuing with the model selection phase, it is imperative to assess whether the explanatory variable being used to construct the ARDL model, which in this case is the first differenced spot price of oil (DOILP), is a useful predictor of the seasonally differenced Canadian unemployment rate (SUER). To do this, a Granger causality test is carried out to see if the spot price of oil successfully predicts the unemployment rate variable. However, it is important to note

[9] The subscript $i{\cdot}s$ and $j{\cdot}s$ in equation 4 represent the lags at which the seasonal AR and MA terms are placed respectively. Since monthly data is used in this study, s=12. So, when i=1 and s=12, $i{\cdot}s$=12, and when i=2 and s=12, $i{\cdot}s$=24. When j=1 and s=12, $j{\cdot}s$=12, and when j=2 and s=12, $j{\cdot}s$=24.

that causality in this situation does not imply causality in the sense that the explantory variable has a causal effect on the dependent variable. Rather, according to Stock and Watson (2012), Granger causality implies that the independent variable at specific lags is a useful predictor of the dependent variable.

The appropriate lag length that is required for the Granger causality test is derived using the lag length selection process for the associated VAR model, and the optimal lag length is chosen based on the AIC criterion. The sample used for the test is the same as the sample used for the estimation of the parameters, which runs from January 1976 to December of 2014. The variables used will not be the variables in levels. Rather, the variables used for the Granger causality test will be the seasonally differenced unemployment rate (SUER) and the nonseasonally differenced oil price (DOILP). After the Granger Causality test has been carried out, if the spot price of oil has been deemed to be an appropriate explanatory variable for the ARDL model the appropriate lag length for each variable for the final model is determined. The appropriate lag length for each variable is determined using the AIC lag length criterion.[10] Once the appropriate number of lags is determined for each of the variables a regression is carried out. The ARDL model in this paper can be expressed as follows:

$$SUER_t = \phi_0 + \sum_{i=1}^{p} \phi_i SUER_{t-p} + \sum_{h=1}^{K} \varphi_h DOILP_{t-h} + \varepsilon_t \quad . \tag{5}$$

After the appropriate lag lengths for the dependent and the independent variables are determined and the estimation has been carried out, the model diagnostic phase is performed in order to assess the stability of the model as well as evaluate whether there is any serial correlation present at the chosen lag lengths. In determining the presence of serial correlation, the preferred test is the Lagrange multiplier (LM) test. If there is serial correlation present, the model is respecified. If the LM test does not indicate serial correlation, the stability of the model is assessed using the

[10]The EVIEWS 10 ARDL model selection process uses the same sample for each estimation. Depending on the maximum number of lags chosen for the dependent variable, EVIEWS will adjust the sample used to select the optimal lag length accordingly. The final estimation, however, will use all of the observations available for the estimation of the parameters. Therefore, unless the selected model is the one that possesses the maximum number of lags, the final model will contain more observations than the sample used to derive the appropriate lag length.

CUSUM test and the CUSUM squares test.

The CUSUM, or the cumulative sum of recursive residuals test,[11] as well as the CUSUM squares test, are two model stability tests developed by Brown et al. (1975). The first test plots the cumulative sum of recursive residuals against time t and assesses whether the plot is outside the 5 percent critical lines. If the plot of the cumulative sum is outside the area between the two 5 percent critical lines throughout the plot then it would be reasonable to conclude that there is coefficient instability. Next, the CUSUM of squares test is carried out. This test is similar to the CUSUM test with the only difference being that movement outside the 5 percent critical lines is indicative of variance instability rather than coefficient instability.

4.3 Forecast Evaluation

After the appropriate ARIMA and ARDL models have been derived, dynamic forecasts of the unemployment rate are carried out and the forecasting accuracy of all of the models used in this paper are assessed. Two forecast horizons are used in this paper: one-step ahead and 3-steps ahead. The initial estimation window runs from January 1976 to December 2014 and is used to forecast the unemployment rate in January 2015 and March 2015. Then an additional observation is added to the sample, and one-step ahead and three-step ahead forecasts are generated for February and April 2015 respectively. This process is continued until the end of the forecast period, December 2018.

The accuracy of the forecasts will be measured by three criteria. Two of these criteria are the root mean squared error (RMSE) and the mean absolute error (MAE). The form for both is expressed below:

$$RMSE : \sqrt{\frac{1}{h}\sum_{T+1}^{T+h}(\widehat{UER}_t - UER_t)^2} \tag{6}$$

[11] According to Früiiwirth-Schnatter (1996) a resursive residual can be defined as the error for the one-step ahead predictive distribution of the future value of the dependent variable which for the ARDL equation in this paper is $SUER_t$.

$$MAE : \sum_{T+1}^{T+h} |\widehat{UER}_t - UER_t| \quad . \tag{7}$$

In the above two equations, $\widehat{UER}_t$ and UER_t represent the forecast and the actual value of the unemployment rate variable in period t, and the value of the forecast is determined for a horizon H period, beginning in period T (Neusser, 2016).

Another measure that is utilized to assess the accuracy of the forecasts is called the mean average percentage error (MAPE). The formula for the MAPE is expressed below in equation 8. According to Kim and Kim (2016), the MAPE is one of the most utilized measures for assessing the accuracy of forecasts because it is "scale-independent" and can be interpreted easily; however, if the actual values of the dependent variable, in this case UER, are very small or less than one, the MAPE produces a very big percentage error, and if the actual values of the dependent variable are zero, then the MAPE produces an infinite value. For this reason, the MAE and RMSE are used in conjunction with MAPE to assess the accuracy of the forecasts.

$$MAPE : \frac{100}{h} \sum_{T+1}^{T+h} |UER_t - \widehat{UER}_t| \quad . \tag{8}$$

After the RMSE, MAE, and MAPE are obtained, the next step involves assessing whether or not the forecasts of each model are biased or unbiased. According to Enders (2014), the bias of the forecasts for each horizon is assessed by estimating the following equation:

$$UER_t = \alpha_0 + \alpha_1 F_t^H + \varepsilon_t^H \quad . \tag{9}$$

For the above equation, t represents the time period, F represents the forecast and H represents the forecast horizon. If the forecast horizon is 1, then H will equal 1 and so on. According to Enders (2014), if $\alpha_0 = 0$ and $\alpha_1 = 1$, then the forecasts are unbiased. If these coefficients are not equal to 0 and 1 respectively, then the forecasts are biased. Bias is assessed by conducting an F test.

Lastly, the final test that is carried out in this paper is a test for predictive accuracy. Although the performance of forecasts can be reliably assessed from the RMSE, MAE, and MAPE, the

Diebold-Mariano (DM) test is required to determine whether the differences in RMSE and MAE are actually significant. According to Diebold (2015), the purpose of the DM test is to examine "the forecast error loss differential" between the forecasts of two models. For instance, suppose a researcher is comparing the forecast errors of two models designated as model 1, with forecast errors $e_{1t} = UER_t - \widehat{UER}_{1t}$, and model 2, with errors $e_{2t} = UER_t - \widehat{UER}_{2t}$. According to Diebold (2015), the loss functions for model 1 and model 2 respectively can be expressed as $L(e_{1t})$ and $L(e_{2t})$, or alternatively as $L(e_{1t}(F_{1t}))$ and $L(e_{2t}(F_{2t}))$, where F_{1t} and F_{2t} are the forecasts generated by models 1 and 2 respectively according to Diebold (2015).

The forecast error loss differential for models 1 and 2 is $d_{12t} = L(e_{1t}(F_{1t})) - L(e_{2t}(F_{2t}))$. The equation for the mean loss differential of h forecasts over the same forecast horizon can be expressed as shown below:

$$\overline{d}_{12} = \frac{1}{h}\sum_{t=1}^{h} d_{12t} \quad . \tag{10}$$

The DM test statistic is based on the mean difference between the loss functions of model 1 and model 2, and it is expressed as shown below:

$$d = \frac{\overline{d}_{12}}{\hat{\sigma}_{\overline{d}_{12}}} \quad . \tag{11}$$

where $\hat{\sigma}_{\overline{d}_{12}}$ represents the estimated standard error associated with $\overline{d}_{12}$. The DM statistic d has a t distribution with $h-1$ degrees of freedom.[12]

The null hypothesis and the alternative hypothesis for the DM test, according to Diebold (2015), can then be expressed as the "zero expected loss differential" hypothesis as shown below:

$$H_0 : E(d_{12t}) = E(L(e_{1t}(F_{1t})) - L(e_{2t}(F_{2t})) = 0 \quad . \tag{12}$$

$$H_A : E(d_{12t}) = E(L(e_{1t}(F_{1t})) - L(e_{2t}(F_{2t})) \neq 0 \quad . \tag{13}$$

[12] According to Diebold (2015) the limiting distribution of d is standard normal, however, both Enders (2014) and EVIEWS 10 treat d as having a t distribution with $h-1$ degrees of freedom.

According to Enders (2014), under the null hypothesis, two alternative loss functions are the squared loss function and absolute loss function. The assumption made when using a squared loss function is that the cost of forecast errors increases with the square of the forecast error, whereas the assumption made when using the absolute loss function is that the cost of forecast errors increases with the magnitude of the errors. The squared loss function puts more emphasis on large forecast errors than the absolute loss function. So, when the test is carried out, both functions are used to assess the accuracy of the forecasts associated with each model. Furthermore, the sign of the test statistic denotes which model is better. If model 1 is better than model 2, the DM statistic will have a negative value, because the mean loss is larger for model 2. If the DM statistic is positive the opposite is true. To successfully reject the null hypothesis, the p-value associated with the DM statistic must be lower than the level of significance.

It is important to note that if the two models being compared are nested, according to Enders (2014), the Diebold-Mariano test will not work because it might not have a t distribution. Also, if the two models are nested, the more parsimonious of the two models, according to Enders (2014), will have fewer errors. Hence, the DM test will be biased towards the simpler model. An alternative test that can be used to assess the predictive accuracy of two nested models is the Clark-West test. When using this test to compare two nested models, according to Enders (2014), model 1 is defined to be the smaller model that is nested within the larger model 2. The larger model, model 2, is the one that contains the greatest number of parameters and the fewest restrictions. Using the same notation as the for DM test elaborated on previously, one first constructs a new variable z_t as follows:

$$z_t = (e_{1t})^2 - [(e_{2t})^2 - (F_{1t} - F_{2t})^2] \qquad t = T, .., H. \tag{14}$$

To carry out the test, z_t is regressed on a constant. According to Enders (2014), under the null hypothesis of equal predictive accuracy of the two nested models, z_t should be zero. To successfully reject the null hypothesis at the 5 percent level, the t statistic for the constant has to be greater than 1.645, since it is a one sided test. If the null hypothesis is rejected in favour of the alternative hypothesis of unequal predictive accuracy, it means that the larger model, model 2, is better than the smaller model, model 1.

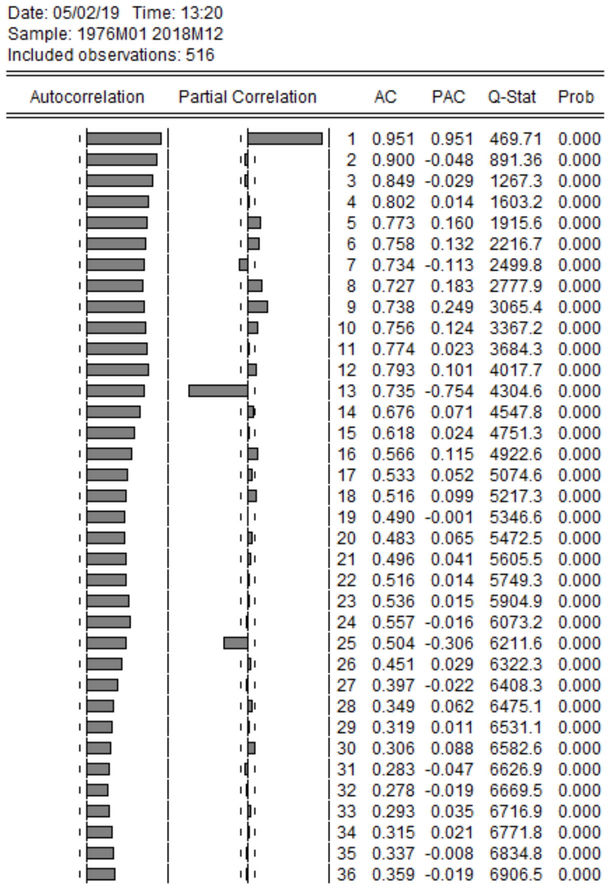
5 Empirical Results

This section of the paper contains the results for the model selection phase for both the ARIMA and ARDL models along with the forecast results for all of the models. The forecast stage consists of carrying out 1-step ahead and 3-steps ahead forecasts of all the models. The performance of the forecasts of each model, as elaborated on previously, will be evaluated based on the RMSE, MAE, and the MAPE. Next, the models will be subjected to a test of bias to determine whether or not the forecasts being produced by the models are unbiased. Finally, the performance of the foreacasts will be further assessed using the Diebold-Mariano (DM) test and the Clark-West test.

5.1 ARIMA Model Selection

5.1.1 Model Identification

The first step in determining the appropriate ARIMA model involves assessing the stationarity of the underlying series. This procedure is carried out in this paper using two steps. The first step involves the qualitative inspection of the correlogram of the unemployment rate in levels shown below in figure 3. The correlogram in figure 3 shows the autocorrelation function (ACF) being very positive and slowly declining. The partial autocorrelation function shows a significant peak at lag 1 but drops off only to show another significant peak at lag 13. There are also a couple of significant smaller spikes in between lag 1 and 13. The series also seems to be exhibiting a seasonal pattern. Overall, figure 3 indicates that the underlying series is nonstationary. However, before proceeding with the model identification step, unit root tests are conducted in order to verify whether the unemployment rate series is nonstationary. It is important to note that for this stage of the model building process, the entire sample is used to determine the stationarity of the series during both the qualitative inspection of the correlogram and the unit root testing.



Date: 05/02/19 Time: 13:20
Sample: 1976M01 2018M12
Included observations: 516

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.951	0.951	469.71	0.000
		2	0.900	-0.048	891.36	0.000
		3	0.849	-0.029	1267.3	0.000
		4	0.802	0.014	1603.2	0.000
		5	0.773	0.160	1915.6	0.000
		6	0.758	0.132	2216.7	0.000
		7	0.734	-0.113	2499.8	0.000
		8	0.727	0.183	2777.9	0.000
		9	0.738	0.249	3065.4	0.000
		10	0.756	0.124	3367.2	0.000
		11	0.774	0.023	3684.3	0.000
		12	0.793	0.101	4017.7	0.000
		13	0.735	-0.754	4304.6	0.000
		14	0.676	0.071	4547.8	0.000
		15	0.618	0.024	4751.3	0.000
		16	0.566	0.115	4922.6	0.000
		17	0.533	0.052	5074.6	0.000
		18	0.516	0.099	5217.3	0.000
		19	0.490	-0.001	5346.6	0.000
		20	0.483	0.065	5472.5	0.000
		21	0.496	0.041	5605.5	0.000
		22	0.516	0.014	5749.3	0.000
		23	0.536	0.015	5904.9	0.000
		24	0.557	-0.016	6073.2	0.000
		25	0.504	-0.306	6211.6	0.000
		26	0.451	0.029	6322.3	0.000
		27	0.397	-0.022	6408.3	0.000
		28	0.349	0.062	6475.1	0.000
		29	0.319	0.011	6531.1	0.000
		30	0.306	0.088	6582.6	0.000
		31	0.283	-0.047	6626.9	0.000
		32	0.278	-0.019	6669.5	0.000
		33	0.293	0.035	6716.9	0.000
		34	0.315	0.021	6771.8	0.000
		35	0.337	-0.008	6834.8	0.000
		36	0.359	-0.019	6906.5	0.000

Figure 3: The correlogram of Canadian unemployment rate in level.

The first unit root test that is carried out is the Augmented Dickey-Fuller test. This is followed by the Dickey-Fuller GLS test and the four M tests (MZt, MZa, MSB, and MPT). Each test was carried out in the presence of an intercept, and then an intercept plus a linear trend. Since Dritsakis and Klazoglou (2018) conducted unit root tests using both an intercept and a trend for the U.S. unemployment rate, this paper follows the same procedure. The unit root test results for the Canadian unemployment rate are shown below in table 2.

As shown in table 2, the results of the ADF test indicate the presence of a unit root in levels because the null hypothesis cannot be rejected at the 1, 5, or 10 percent levels of significance. However, when the series is differenced once it is made stationary as evidenced by the rejection of the null hypothesis at the 5 percent level of significance. The results are the same when the test is carried out with an intercept as well as an intercept plus a trend.

Table 2: The results of the unit root tests. Null Hypothesis: UER has a unit root.

	Intercept		Intercept & Trend	
Unit Root Test	Lag Length	Test Statistic	Lag Length	Test Statistic
ADF				
Levels	17	-2.089532	17	-3.084661
1st Diff	11	-3.426505 ∗∗	11	-3.435216 ∗∗
DF-GLS				
Levels	17	-2.108993 ∗∗	17	-2.338224
1st Diff			11	-3.078647 ∗∗
MZt				
Levels	17	-11.582 ∗∗	17	-13.4863
1st Diff			11	-0.97732
Mza				
Levels	17	-2.27871 ∗∗	17	-2.54384
1st Diff			11	-0.69873
MSB				
Levels	17	0.19675 ∗∗	17	0.18862
1st Diff			11	0.71494
MPT				
Levels	17	2.61874 ∗∗	17	7.06808
1st Diff			11	93.1690

∗ significant at $p < 0.10$, ∗∗ significant at $p < 0.05$, ∗∗∗ significant at $p < 0.01$

The next test that is carried out is the Dickey-Fuller GLS test. As shown in table 2, when the test is carried in the presence of only an intercept, the null hypothesis can be successfully rejected at the 5 percent level. This indicates that the series does not have a unit root which suggests that the unemployment rate series is stationary. However, when DF-GLS test is carried out in the presence of an intercept plus a trend, the null hypothesis cannot be rejected at any level of significance. In this case, the null hypothesis is rejected at the 5 percent level of significance only after differencing the series once.

The next unit root tests that are examined in order to assess the sationarity of the unemployment rate series are the four M tests. As shown in table 2, all of the 4 test statistics reject the null hypothesis of a unit root at the 5 percent level of significance in the presence of only an intercept, which suggests that the unemployment rate series is stationary. However, when the four M tests are conducted using both an intercept and a trend, the null hypothesis is not rejected at any level of significance even after differencing the unemployment rate series once.

Based on the results of all of the nonseasonal unit root tests, it cannot be concluded with a high degree of confidence that the unemployment rate series is a nonstationary series. The DF-GLS test as well as the four M tests reveal that the null hypothesis cannot successfully be rejected when the unit root tests are carried out using an intercept plus a trend. It can only be successfully rejected if an intercept is used without the trend. The ADF test results are similar to results obtained in other papers which forecasted the unemployment rate. However, what is important to keep in mind is that stationarity can rarely be quickly ascertained based on the results of just one unit root test. Most other papers in the literature often use at least two unit root tests in order to make a comparison between the tests and make a judgment based on their results.

There are instances in which one test yields results that lead to the rejection of the null hypothesis of the existence of a unit root while the other test fails to reject the null hypothesis. For this reason, a better conclusion can be reached if two separate unit root tests achieve similar results. If they don't reach the same conclusion then stationarity cannot be assessed with a high degree of confidence. In this paper, not all of the unit root tests indicate that the series is nonstationary. However, based on the results of the some unit root tests along with the qualitative inspection of the correlogram shown in figure 3, it can be concluded with a reasonable degree of certainty that the unemployment rate series is a nonstationary series. So it must be converted into a stationary series before it can be modeled. To convert the series into a stationary series, the appropriate level of differencing must be determined.

The correlogram shown in figure 3 shows the unemployment rate series exhibiting clear seasonal patterns. For this reason, it seems prudent to evaluate the seasonality of the series before deciding on the appropriate degree of differencing. The ACF in particular exhibits wave like fluctuations which is also indicative of a clear seasonal pattern. This observation suggests that there could be a seasonal unit root at specific frequencies. To verify the existance of a seasonal unit root in the unemployment rate series, the HEGY seasonal unit root test is conducted. The results are shown below in table 3. According to the results, when the HEGY test is carried out with only an intercept, there is a unit root at the zero frequency (which denotes the nonseasonal unit root for this test), the 4 months per cycle frequency, and the 3 months per cycle frequency. When the test is carried out in the presence of both an intercept and a trend, there is a unit root at only 4 months per cycle and 3 months per cycle frequencies. Based on these observations, it is not entirely clear whether there is in fact a seasonal unit root since the null hypothesis can be successfully rejected at most of the frequencies. However, because a seasonal unit root is reported for at least two of the frequencies, seasonal differencing was used to transform the series into seasonally differenced series denoted by the variable SUER.

To determine whether or not SUER is a stationary series, all of the unit root tests, nonseasonal unit root tests as well as the HEGY test, are repeated. As shown in table 3 and table 4, after the series is subjected to seasonal differencing, all unit root tests indicate that the unemployment rate series has now become a stationary series. It is important to note that the four M tests in particular resulted in a rejection of the null hypothesis of a unit root even after the unit root test is conducted using an intercept and a trend as shown in table 4. This is in contrast to the results in table 2 which show that when a unit root test is conducted on the unemployment rate series using an intercept and a trend, the null hypothesis can not be rejected even after subjecting the series to first differencing. Therefore, based on the results in table 3 and table 4, it can be concluded with a fair degree of confidence that the appropriate level of integration for the monthly Canadian unemployment rate series is one seasonal difference.

Table 3: HEGY Seasonal Unit Root Test for the Canadian unemployment rate.

Null Hypothesis		Intercept Lag Length	Intercept Test Statistic	Intercept and Trend Lag Length	Intercept and Trend Test Statistic
Seasonal unit root (Zero frequencey)					
	Levels	5	-2.3068	5	-3.2876^{*}
	Seasonal Diff	6	-4.52^{***}	6	-4.56^{***}
Seasonal unit root (2 months per cycle)					
	Levels	5	-1.5603	5	-1.551
	Seasonal Diff	6	-7.79^{***}	6	-7.78^{***}
Seasonal unit root (4 months per cycle)					
	Levels	5	7.67^{***}	5	7.599^{***}
	Seasonal Diff	6	44.5^{***}	6	44.4^{***}
Seasonal unit root (2.4 months per cycle)					
	Levels	5	2.2883^{*}	5	2.31^{*}
	Seasonal Diff	6	46.83^{***}	6	46.71^{***}
Seasonal unit root (12 months per cycle)					
	Levels	5	3.1674^{**}	5	3.14^{**}
	Seasonal Diff	6	68.90^{***}	6	68.85^{***}
Seasonal unit root (3 months per cycle)					
	Levels	5	2.8115^{*}	5	2.833^{*}
	Seasonal Diff	6	58.13^{***}	6	58.01^{***}
Seasonal unit root (6 months per cycle)					
	Levels	5	1.8953	5	1.9577
	Seasonal Diff	6	59.17^{***}	6	59.15^{***}

$*$ significant at $p < 0.10$, $**$ significant at $p < 0.05$, $***$ significant at $p < 0.01$

Table 4: The results of the unit root tests. Null Hypothesis: UER has a unit root.

	Intercept		Intercept & Trend	
Unit Root Test	Lag Length	Test Statistic	Lag Length	Test Statistic
ADF				
Seasonal Diff	12	-3.359919 $**$	12	-3.388993 $*$
DF-GLS				
Seasonal Diff	12	-2.769366 $***$	12	-3.324586 $**$
MZt				
Seasonal Diff	12	-16.1152 $***$	12	-23.5066 $**$
Mza				
Seasonal Diff	12	-2.82897 $***$	12	-3.42626 $***$
MSB				
Seasonal Diff	12	0.17555 $**$	12	0.14576 $**$
MPT				
Seasonal Diff	12	1.55729 $***$	12	3.88925 $***$

$*$ significant at $p < 0.10$, $**$ significant at $p < 0.05$, $***$ significant at $p < 0.01$

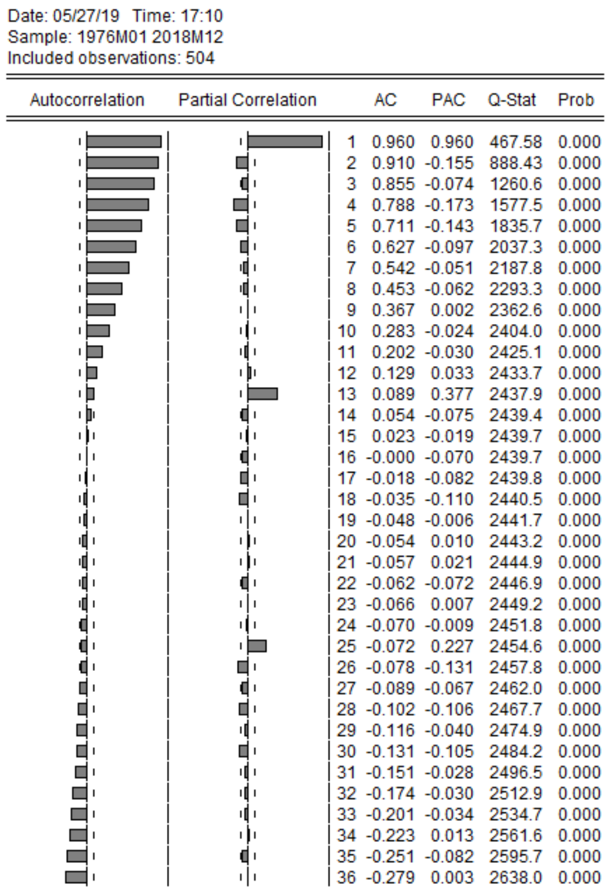
If the only transformation the Canadian unemployment rate series is subjected to is one seasonal differencing, it can be modelled as a seasonal random walk.[13] However, an ARIMA model that will include the appropriate AR(p) or MA(q) terms will be used to construct the final univariate model. The first step towards determining these terms starts with an evaluation of the correlogram of the seasonally differenced series shown below in figure 4.

The ACF of the correlogram in figure 4 gradually declines towards zero while the PACF has a large positive spike at lag 1 then cuts off to zero only to have a few small negative spikes and then one small positive spike at lag 13. The pattern observed in figure 4 appears to be a pattern which is exhibited by an AR(p) process. To model this series, several seasonal ARIMA models with varying AR(p) processes are estimated, starting with an AR(1) process all the way to an AR(12) process. The model that is selected will be the model that is associated with the lowest AIC and SBIC as well as the greatest number of coefficients significant at the 10 percent level.

If two models have very similar AIC and SBIC values as well as the same number of significant coefficients, the model that will be selected will be the one that is the most parsimonious model. As mentioned previously, a parsimonious model according to Ledolter and Abraham (1981) is a model containing "the smallest possible number of parameters." Therefore, the other lags shown in the correlogram will not be taken into account because they will not contribute to the development of a parsimonious model.

Since the sample used for the estimation of the parameters in this paper is from January 1976 to December 2014, and the maximum number of AR terms that will be evaluated during the ARIMA model selection process is 12, the sample that will be used in all of these estimations will range from January 1978 to December 2014. However, the final estimation will not use the same sample range. In fact, the sample range for the final model will be higher than the range chosen for the model identification process.

[13]The seasonal random walk model for the 1 step ahead forecast is obtained by transforming the unemployment rate series as follows: srwf1 = $UER_{t-1} + UER_{t-12} - UER_{t-13}$. For the 3-steps ahead forecast the seasonal random walk model is obtained by transforming the unemployment rate series as follows: $UER_{t-3} + UER_{t-12} - UER_{t-15}$. The 1-step and 3-steps ahead forecasts are carried out using sample ranges January 2015 to December 2018 and March 2015 to December 2018 respectively.



Date: 05/27/19 Time: 17:10
Sample: 1976M01 2018M12
Included observations: 504

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.960	0.960	467.58	0.000
		2	0.910	-0.155	888.43	0.000
		3	0.855	-0.074	1260.6	0.000
		4	0.788	-0.173	1577.5	0.000
		5	0.711	-0.143	1835.7	0.000
		6	0.627	-0.097	2037.3	0.000
		7	0.542	-0.051	2187.8	0.000
		8	0.453	-0.062	2293.3	0.000
		9	0.367	0.002	2362.6	0.000
		10	0.283	-0.024	2404.0	0.000
		11	0.202	-0.030	2425.1	0.000
		12	0.129	0.033	2433.7	0.000
		13	0.089	0.377	2437.9	0.000
		14	0.054	-0.075	2439.4	0.000
		15	0.023	-0.019	2439.7	0.000
		16	-0.000	-0.070	2439.7	0.000
		17	-0.018	-0.082	2439.8	0.000
		18	-0.035	-0.110	2440.5	0.000
		19	-0.048	-0.006	2441.7	0.000
		20	-0.054	0.010	2443.2	0.000
		21	-0.057	0.021	2444.9	0.000
		22	-0.062	-0.072	2446.9	0.000
		23	-0.066	0.007	2449.2	0.000
		24	-0.070	-0.009	2451.8	0.000
		25	-0.072	0.227	2454.6	0.000
		26	-0.078	-0.131	2457.8	0.000
		27	-0.089	-0.067	2462.0	0.000
		28	-0.102	-0.106	2467.7	0.000
		29	-0.116	-0.040	2474.9	0.000
		30	-0.131	-0.105	2484.2	0.000
		31	-0.151	-0.028	2496.5	0.000
		32	-0.174	-0.030	2512.9	0.000
		33	-0.201	-0.034	2534.7	0.000
		34	-0.223	0.013	2561.6	0.000
		35	-0.251	-0.082	2595.7	0.000
		36	-0.279	0.003	2638.0	0.000

Figure 4: The correlogram of Canadian unemployment rate in seasonal difference.

Table 5: Seasonal ARIMA models with varying AR(p) terms.

Model	AIC	SBIC	Significant Coefficients	Adjusted Sample
$\text{ARIMA}(1,0,0)(0,1,0)_{12}$	0.471428	0.489878	1	1978M01 to 2014M12
$\text{ARIMA}(2,0,0)(0,1,0)_{12}$	0.447258	0.474932	2	1978M01 to 2014M12
$\text{ARIMA}(3,0,0)(0,1,0)_{12}$	0.446907	0.483806	1	1978M01 to 2014M12
$\text{ARIMA}(4,0,0)(0,1,0)_{12}$	0.415526	0.461651	3	1978M01 to 2014M12
$\textbf{ARIMA}(5,0,0)(0,1,0)_{12}$	0.394661	**0.450010**	3	1978M01 to 2014M12
$\textbf{ARIMA}(6,0,0)(0,1,0)_{12}$	**0.390801**	0.455375	3	1978M01 to 2014M12
$\text{ARIMA}(7,0,0)(0,1,0)_{12}$	0.392630	0.466429	2	1978M01 to 2014M12
$\text{ARIMA}(8,0,0)(0,1,0)_{12}$	0.393508	0.476532	2	1978M01 to 2014M12
$\text{ARIMA}(9,0,0)(0,1,0)_{12}$	0.397792	0.490041	2	1978M01 to 2014M12
$\text{ARIMA}(10,0,0)(0,1,0)_{12}$	0.401998	0.503471	2	1978M01 to 2014M12
$\text{ARIMA}(11,0,0)(0,1,0)_{12}$	0.405932	0.516630	2	1978M01 to 2014M12
$\text{ARIMA}(12,0,0)(0,1,0)_{12}$	0.408482	0.528405	2	1978M01 to 2014M12

As shown in table 5 the most appropriate model for this series based on the lowest AIC criterion is the AR(6) seasonal ARIMA model, which can be expressed as an ARIMA $(6,0,0)(0,1,0)_{12}$. However, the most appropriate model based on the SBIC is the AR(5) seasonal ARIMA model, which can be expressed as $(5,0,0)(0,1,0)_{12}$. It is important to note that both of these models contain the same number of statistically significant coefficients. So, the only way to decide between these two models is to use the concept of model parsimony. Since the AR(5) model has fewer parameters than the AR(6) model, the AR(5) model will be chosen as the model that will be evaluated in the model diagnostic step.

5.1.2 Model Diagnostics

In this section of the paper, the model chosen during the model identification stage, the AR(5) seasonal ARIMA model, will be estimated using the sample range of June 1977 to December 2014 and the correlogram of the residuals will be evaluated to determine if the ACF and PACF plots contain significant spikes at different lag lengths. According to figure 5, there is one significant spike observed at lag 12 in the ACF, and spikes observed at lags 12, 24, and 36 in the PACF. These results indicate that there is additional seasonality that needs to be captured by adding a seasonal MA term at lag 12.[14] After the seasonal MA(1) term is added to the AR(5) seasonal ARIMA model it can now be expressed as an ARIMA $(5,0,0)(0,1,1)_{12}$. After the new model is estimated, the ACF and PACF of the residuals are examined.

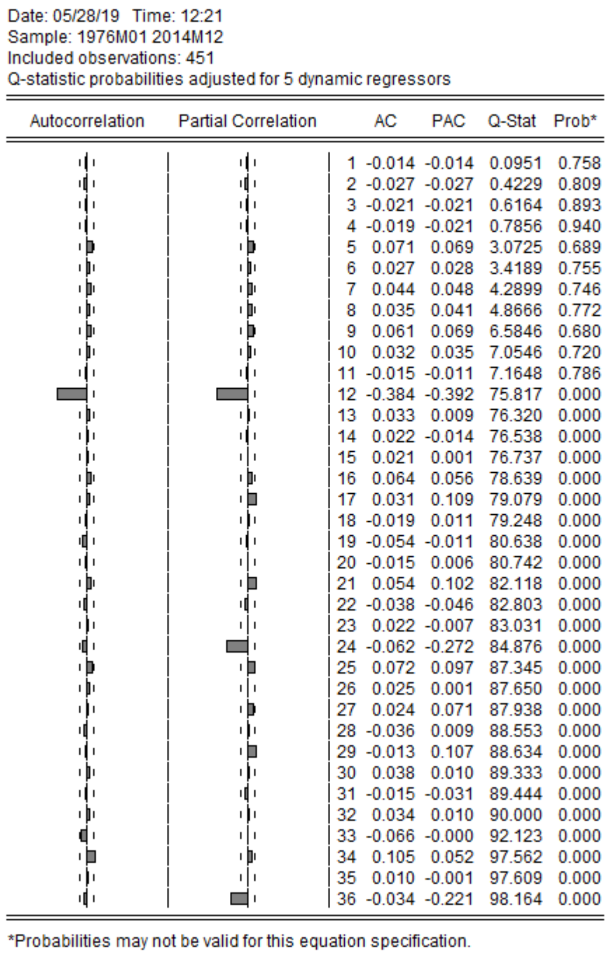
As shown in figure 6, the additional seasonal moving average term turns the residuals into white noise. It is now possible to proceed with the forecasting stage using this model. However, although the AR(5) seasonal ARIMA model can be used to conduct forecasts, not all of its coefficients are significant. An alternative model that can be used to forecast the unemployment rate is an ARIMA model that contains only the statistically significant AR terms from the AR(5) seasonal ARIMA model. To determine which model has the better fit, the two models are estimated and the best model is chosen based on which model achieves the lowest AIC and SBIC.

[14]The full expression for the ARIMA model is $(p,d,q)(P,D,Q)_{12}$. So, p denotes AR terms, q denotes MA terms, P denotes seasonal AR terms, and Q denotes seasonal MA terms.

The model chosen during the model identification stage clearly contains coefficients that are not statistically significant, as shown in table 6. Therefore, the terms whose coefficients are not statistically significant are removed, and a new model containing only the terms with significant coefficients is estimated. According to the results shown in table 6, the best candidate for the benchmark seasonal ARIMA model is the ARIMA $(||1,3,5||,0,0)(0,1,1)_{12}$ because it achieves the lowest AIC and BIC.[15] Furthermore, as shown in figure 7, the correlogram of its residuals exhibits white noise behavior. Therefore, this restricted seasonal ARIMA model will be used to conduct forecasts and its performance will be compared to those of other models.

As shown in table 6, the coefficient α_1 in the ARIMA $(||1,3,5||,0,0)(0,1,1)_{12}$ is almost identical to 1, which is consistent with there also being a unit root. Although the existence of a trend is not supported by the unit root tests, the unemployment rate series as shown in figure 1 does seem to show a trend, which necessitates subjecting the unemployment rate series to a non-seasonal difference. However, since the Canadian unemployment rate exhibits strong seasonality as shown by the unit root tests, it does not make sense to model the series as having only a nonseasonal difference. So, for this paper, an additional ARIMA model will be estimated in which the unemployment rate is subjected to both nonseasonal and seasonal differencing. The same model building process that was used for the previous ARIMA is also used to construct this new ARIMA model. The final model obtained is an ARIMA $(||3,4,5||,1,0)(0,1,1)_{12}$.

[15]The notation "||...||" around the AR terms represents the restricted model containing only the statistically significant coefficients.



Date: 05/28/19 Time: 12:21
Sample: 1976M01 2014M12
Included observations: 451
Q-statistic probabilities adjusted for 5 dynamic regressors

Lag	AC	PAC	Q-Stat	Prob*
1	-0.014	-0.014	0.0951	0.758
2	-0.027	-0.027	0.4229	0.809
3	-0.021	-0.021	0.6164	0.893
4	-0.019	-0.021	0.7856	0.940
5	0.071	0.069	3.0725	0.689
6	0.027	0.028	3.4189	0.755
7	0.044	0.048	4.2899	0.746
8	0.035	0.041	4.8666	0.772
9	0.061	0.069	6.5846	0.680
10	0.032	0.035	7.0546	0.720
11	-0.015	-0.011	7.1648	0.786
12	-0.384	-0.392	75.817	0.000
13	0.033	0.009	76.320	0.000
14	0.022	-0.014	76.538	0.000
15	0.021	0.001	76.737	0.000
16	0.064	0.056	78.639	0.000
17	0.031	0.109	79.079	0.000
18	-0.019	0.011	79.248	0.000
19	-0.054	-0.011	80.638	0.000
20	-0.015	0.006	80.742	0.000
21	0.054	0.102	82.118	0.000
22	-0.038	-0.046	82.803	0.000
23	0.022	-0.007	83.031	0.000
24	-0.062	-0.272	84.876	0.000
25	0.072	0.097	87.345	0.000
26	0.025	0.001	87.650	0.000
27	0.024	0.071	87.938	0.000
28	-0.036	0.009	88.553	0.000
29	-0.013	0.107	88.634	0.000
30	0.038	0.010	89.333	0.000
31	-0.015	-0.031	89.444	0.000
32	0.034	0.010	90.000	0.000
33	-0.066	-0.000	92.123	0.000
34	0.105	0.052	97.562	0.000
35	0.010	-0.001	97.609	0.000
36	-0.034	-0.221	98.164	0.000

*Probabilities may not be valid for this equation specification.

Figure 5: The correlogram of the residuals for the ARIMA$(5,0,0)(0,1,0)_{12}$.

As shown below in table 6, the first column contains the ARIMA $(5,0,0)(0,1,1)_{12}$ model, the second column contains the ARIMA $(||1,3,5||,0,0)(0,1,1)_{12}$ model, and last column contains the ARIMA $(||3,4,5||,1,0)(0,1,1)_{12}$ model. Throughout the rest of this paper, the first ARIMA model, in which there is only one seasonal differencing of the unemployment rate, will be referred to as the SARIMA model, while the second ARIMA model, which also includes non-seasonal differencing, will be referred to as the TARIMA model.

Table 6: The estimation of the final ARIMA models.

	ARIMA (5,0,0)(0,1 ,1)12	SARIMA p = 1,3,5 SMA(12)	TARIMA p = 3,4,5 SMA(12)
α_0	-0.002		
	(0.004)		
α_1	1.064***	1.024***	
	(0.046)	(0.028)	
α_2	-0.078		
	(0.069)		
α_3	0.162**	0.099**	0.146***
	(0.067)	(0.043)	(0.043)
α_4	-0.052		0.105**
	(0.074)		(0.047)
α_5	-0.122***	-0.150***	-0.133**
	(0.049)	(0.029)	(0.047)
Θ_{12}	0.060***	-0.695***	-0.700***
	(0.037)	(0.036)	(0.036)
AIC	0.072	0.063	0.064
SBIC	0.145	0.109	0.11
Adjusted Sample	1977M06 to 2014M12	1977M06 to 2014M12	1977M06 to 2014M12

Standard errors are in parentheses. * significant at $p < 0.10$,
** significant at $p < 0.05$, *** significant at $p < 0.01$

Date: 05/28/19 Time: 13:21
Sample: 1976M01 2014M12
Included observations: 451
Q-statistic probabilities adjusted for 1 ARMA term and 5 dynamic regressors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.017	-0.017	0.1236	
		2	-0.031	-0.032	0.5723	0.449
		3	-0.008	-0.009	0.6003	0.741
		4	-0.002	-0.003	0.6016	0.896
		5	0.099	0.099	5.1164	0.276
		6	0.018	0.021	5.2631	0.385
		7	0.039	0.046	5.9459	0.429
		8	-0.017	-0.013	6.0862	0.530
		9	0.064	0.067	7.9726	0.436
		10	-0.032	-0.041	8.4559	0.489
		11	-0.052	-0.054	9.7292	0.465
		12	0.037	0.025	10.360	0.498
		13	-0.004	-0.005	10.366	0.584
		14	-0.030	-0.044	10.776	0.630
		15	0.006	0.011	10.793	0.702
		16	0.023	0.029	11.046	0.749
		17	0.062	0.065	12.827	0.685
		18	0.006	0.008	12.843	0.747
		19	-0.060	-0.048	14.533	0.694
		20	-0.039	-0.034	15.270	0.705
		21	0.069	0.055	17.511	0.620
		22	-0.011	-0.029	17.565	0.676
		23	0.011	0.015	17.627	0.728
		24	-0.048	-0.049	18.733	0.717
		25	0.014	0.020	18.829	0.761
		26	0.000	-0.012	18.829	0.805
		27	0.011	0.020	18.883	0.841
		28	-0.022	-0.020	19.112	0.866
		29	0.002	0.016	19.114	0.895
		30	0.054	0.037	20.551	0.875
		31	-0.038	-0.022	21.265	0.879
		32	-0.026	-0.028	21.582	0.896
		33	-0.029	-0.033	21.989	0.908
		34	0.080	0.071	25.131	0.835
		35	0.022	0.019	25.379	0.857
		36	-0.020	-0.006	25.582	0.878

*Probabilities may not be valid for this equation specification.

Figure 6: The correlogram of the residuals for the ARIMA $(5,0,0)(0,1,1)_{12}$.

Date: 05/28/19 Time: 17:24
Sample: 1976M01 2014M12
Included observations: 451
Q-statistic probabilities adjusted for 1 ARMA term and 3 dynamic regressors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.018	0.018	0.1462	
		2	-0.064	-0.065	2.0271	0.155
		3	0.008	0.011	2.0594	0.357
		4	-0.016	-0.021	2.1828	0.535
		5	0.094	0.097	6.2635	0.180
		6	0.025	0.019	6.5606	0.255
		7	0.033	0.046	7.0583	0.315
		8	-0.015	-0.017	7.1635	0.412
		9	0.061	0.071	8.8891	0.352
		10	-0.032	-0.047	9.3580	0.405
		11	-0.056	-0.048	10.834	0.371
		12	0.036	0.022	11.422	0.409
		13	-0.004	-0.007	11.429	0.493
		14	-0.031	-0.041	11.870	0.538
		15	0.007	0.011	11.892	0.615
		16	0.026	0.030	12.210	0.663
		17	0.062	0.066	14.008	0.598
		18	0.009	0.007	14.045	0.664
		19	-0.065	-0.050	16.044	0.589
		20	-0.036	-0.028	16.669	0.612
		21	0.066	0.053	18.760	0.537
		22	-0.006	-0.028	18.775	0.600
		23	0.007	0.015	18.798	0.658
		24	-0.045	-0.052	19.783	0.655
		25	0.011	0.024	19.836	0.706
		26	0.003	-0.014	19.841	0.755
		27	0.010	0.023	19.892	0.797
		28	-0.022	-0.023	20.121	0.826
		29	0.005	0.022	20.131	0.860
		30	0.053	0.033	21.511	0.840
		31	-0.035	-0.020	22.111	0.850
		32	-0.030	-0.029	22.543	0.865
		33	-0.023	-0.027	22.793	0.885
		34	0.078	0.071	25.773	0.811
		35	0.031	0.021	26.260	0.826
		36	-0.028	-0.012	26.642	0.844

*Probabilities may not be valid for this equation specification.

Figure 7: The correlogram of the residuals for the ARIMA $(||1,3,5||,0,0)(0,1,1)_{12}$.

Date: 08/30/19 Time: 17:48
Sample: 1976M01 2014M12
Included observations: 450
Q-statistic probabilities adjusted for 1 ARMA term and 3 dynamic regressors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.065	0.065	1.9398	
		2	-0.020	-0.024	2.1202	0.145
		3	0.003	0.006	2.1236	0.346
		4	0.012	0.011	2.1859	0.535
		5	0.009	0.008	2.2268	0.694
		6	0.020	0.019	2.4090	0.790
		7	0.029	0.027	2.7989	0.834
		8	-0.038	-0.041	3.4599	0.839
		9	0.039	0.045	4.1588	0.843
		10	-0.056	-0.065	5.6078	0.778
		11	-0.071	-0.062	7.9135	0.637
		12	0.014	0.020	8.0019	0.713
		13	-0.020	-0.027	8.1914	0.770
		14	-0.045	-0.040	9.1385	0.762
		15	-0.002	0.006	9.1406	0.822
		16	0.014	0.010	9.2271	0.865
		17	0.049	0.058	10.335	0.849
		18	-0.005	-0.014	10.348	0.888
		19	-0.063	-0.060	12.197	0.837
		20	-0.052	-0.038	13.457	0.814
		21	0.056	0.051	14.962	0.779
		22	-0.019	-0.035	15.136	0.816
		23	0.005	0.014	15.146	0.856
		24	-0.039	-0.052	15.884	0.860
		25	0.003	0.012	15.890	0.892
		26	-0.006	-0.005	15.905	0.918
		27	0.013	0.017	15.984	0.937
		28	-0.018	-0.018	16.138	0.950
		29	0.001	0.007	16.138	0.964
		30	0.051	0.035	17.399	0.956
		31	-0.032	-0.027	17.911	0.960
		32	-0.019	-0.017	18.087	0.968
		33	-0.018	-0.026	18.248	0.975
		34	0.084	0.080	21.695	0.934
		35	0.025	0.016	21.991	0.944
		36	-0.016	-0.013	22.111	0.956

*Probabilities may not be valid for this equation specification.

Figure 8: The correlogram of the residuals for the ARIMA $(||3,4,5||,1,0)(0,1,1)_{12}$.

It is important to note that the correlogram of the residuals for the ARIMA $(||3,4,5||,1,0)(0,1,1)_{12}$ shown in figure 8, and the ARIMA $(||1,3,5||,0,0)(0,1,1)_{12}$ shown in figure 7 both show the respective residuals exhibiting white noise behaviour. Although table 6 shows that the SARIMA model hasa slightly lower AIC and SBIC than the TARIMA model, the difference is not that great. Therefore, both models can now be used to conducted forecasts.

5.2 ARDL Model Selection

5.2.1 Model Identification

The section contains the results of the ARDL model identification stage. The first step taken towards the construction of the ARDL model requires the assessment of the stationarity of the dependent variable and the independent variable. The Canadian unemployment rate has already been determined to possess a seasonal unit root which requires one degree of seasonal differencing to make it stationary. Figure 2 suggests that the oil price series in this paper appears to have a clear trend; however, it does not appear to exhibit clear seasonal patterns. Therefore, a seasonal unit root test is not necessary for assessing the stationarity of this series. Therefore, the nonseasonal unit root tests are carried out in order to assess the stationarity of the oil price series.

Table 7: The results of the unit root tests. Null Hypothesis: oilp has a unit root.

	Intercept		Intercept & Trend	
Unit Root Test	Lag Length	Test Statistic	Lag Length	Test Statistic
ADF				
Levels	6	-1.925133	6	-2.422379
1st Diff	1	-11.6483 $***$	1	-11.6392 $***$
DF-GLS				
Levels	6	-1.110499	6	-2.45212
1st Diff	1	-11.36524 $***$	1	-11.19502 $***$
MZa				
Levels	6	-3.00484	6	-13.1366
1st Diff	1	-188.641 $***$	1	-182.146 $***$
MZt				
Levels	6	-1.14326	6	-2.52819
1st Diff	1	-9.65514 $***$	1	-9.51528 $***$
MSB				
Levels	6	0.38047	6	0.19245
1st Diff	1	0.05118 $***$	1	0.05224 $***$
MPT				
Levels	6	8.00175	6	7.13927
1st Diff	1	0.20976 $***$	1	0.58197 $***$

$*$ significant at $p < 0.10$, $**$ significant at $p < 0.05$, $***$ significant at $p < 0.01$

Based on the results in table 8, it appears that the oil price series has a unit root and in order to make it stationary, the series is subjected to one nonseasonal differencing. Thus, the dependent variable in the ARDL model will be the seasonally differenced Canadian unemployment rate, while the exogenous variable will be the nonseasonally differenced oil price.

To determine whether the oil price is a useful predictor of the Canadian unemployment rate, a Granger Causality test is performed. The first step in conducting the Granger Causality test is to determine the appropriate number of lags to use for both variables being examined. The lag length selection was carried out using the VAR lag selection process and the criterion used for this assessment is the AIC. The optimal lag length chosen for the VAR model is 6 lags. The results for the Granger Causality test are shown in table 9. According to the test results, there is only a one-way Granger causal effect between the unemployment rate and the oil price. The oil price Granger causes the unemployment rate but the reverse is not true. What this means is that the spot oil price contains information that is useful for predicting changes in the Canadian unemployment rate.

Table 8: VAR Granger Causality/Block Exogeneity Wald Tests

Sample: 1976M01 2014M12
Included observations: 450

Dependent variable: SUER

Excluded	Chi-sq	df	Prob.
DOILP	14.64684	6	0.0232

Dependent variable: DOILP

Excluded	Chi-sq	df	Prob.
SUER	7.821539	6	0.2515

After the stationarity of both variables is attained and the spot price of oil has been determined to Granger cause the Canadian unemployment rate, the appropriate number of lags for each variable required to build the ARDL model is determined using the AIC lag length criterion.[16] According to the AIC criterion, the appropriate number of lags for the seasonally differenced unemployment rate is 6 while for the nonseasonally differenced oil price it is 4. Hence, the final ARDL model is determined to be an ARDL (6,4) model. Table 10 below shows the results for the ARDL(6,4) estimation.

[16]Since the observations used for the estimation of the parameters in this paper are January 1976 to December 2014, and the maximum number of lags was set to 12 for both the dependent and independent variable in the ARDL selection procedure in EVIEWS, the sample that is used for the model selection process is January 1978 to December of 2014. However, the final sample that is used for estimating the chosen ARDL model is July 1977 to December 2014.

Table 9: Regression result for the ARDL(6,4) with the seasonally differenced Canadian unemployment rate as the dependent variable and the first differenced spot oil price as the exogenous variable.

Constant	0.00008
	(0.014)
$SUER_{t-1}$	1.0457^{***}
	(0.047)
$SUER_{t-2}$	-0.082
	(0.069)
$SUER_{t-3}$	0.1199^{*}
	(0.0694)
$SUER_{t-4}$	0.0034
	(0.0693)
$SUER_{t-5}$	-0.0683
	(0.0699)
$SUER_{t-6}$	-0.0922^{*}
	(0.047)
$DOILP_{t}$	0.00285
	(0.004)
$DOILP_{t-1}$	-0.0097^{*}
	(0.0043)
$DOILP_{t-2}$	0.005
	(0.0044)
$DOILP_{t-3}$	-0.006
	(0.0044)
$DOILP_{t-4}$	-0.0072^{*}
	(0.004)
R-squared	0.934
Adjusted	1977 M07 to
Sample	2014 M12

Standard errors are in parentheses. $*$ significant at $p < 0.10$,
$**$ significant at $p < 0.05$, $***$ significant at $p < 0.01$

The process that produced the results shown in table 9 is repeated using the twice differenced uenmployment rate and the nonseasonally differenced price of oil becasue once again the coefficient of the lagged first difference is close to 1. The ARDL model building process yielded an ARDL(12,4) model when the twice differenced unemployment rate was used. After dropping the coefficients that are not statistically significant for both models, the models obtained are ARDL($||1,6||,||4||$) and ARDL($||3,4,12||,||3||$) respectively. For the rest of this paper, the ARDL($||1,6||,||4||$) model will be referred to as the SARDL model and the ARDL($||3,4,12||,||3||$)

model is referred to as the TARDL model. The estimations for both the SARDL and TARDL models are shown in tables 10 and 11 respectively.[17]

Table 10: Regression result for the ARDL($||1,6||,||4||$) with the seasonally differenced Canadian unemployment rate as the dependent variable and the first differenced spot oil price as the exogenous variable.

$SUER_{t-1}$	1.039^{***}
	(0.018)
$SUER_{t-6}$	-0.112^{***}
	(0.018)
$DOILP_{t-4}$	-0.009^{**}
	(0.004)
R-squared	0.934
Adjusted	1977 M07 to
Sample	2014 M12

Standard errors are in parentheses. $*$ significant at p < 0.10,
$**$ significant at p < 0.05, $***$ significant at p < 0.01

Table 11: Regression result for the ARDL($||3,4,12||,||3||$) with the seasonally and nonseasonally differenced Canadian unemployment rate as the dependent variable and the first differenced spot oil price as the exogenous variable.

$TUER_{t-3}$	0.129^{***}
	(0.042)
$TUER_{t-4}$	0.120^{***}
	(0.042)
$TUER_{t-12}$	-0.416^{***}
	(0.042)
$DOILP_{t-3}$	-0.009^{***}
	(0.003)
R-squared	0.934
Adjusted	1977 M07 to
Sample	2014 M12

Standard errors are in parentheses. $*$ significant at p < 0.10,
$**$ significant at p < 0.05, $***$ significant at p < 0.01

5.2.2 Model diagnostics

The first step of the model diagnostic stage involves the assessment of serial correlation in the chosen ARDL model. The test of choice in this matter is the Breusch-Godfrey Lagrange multiplier

[17] As stated previously, $SUER$ stands for the seasonally differenced unemployment rate, while $TUER$ stands for the seasonally and nonseasonally differenced unemployment rate.

(LM) test. Since the appropriate lag lengths determined for the seasonally differenced unemployment rate and the nonseasonally differenced oil price is 6 and 4 respectively, the lag lengths chosen for the LM test are between 1 and 6. The goal of this process is to determine whether or not there is serial correlation at any lags between 1 and 6.

Based on the results shown in table 11, the null hypothesis of no serial correlation cannot be rejected at any of the chosen lag orders. Therefore, it can be concluded that there is no serial correlation at lags 1 through lag 6 in the SARDL model. At this stage of the model selection process, it reasonable to conclude that the ARDL model, with the spot price of oil, contains an appropriate explanatory variable according to the results of the Granger causality test, and it does not exhibit any serial correlation.

Table 12: Breusch-Godfrey Serial Correlation LM test, SARDL. Null Hypothesis: No serial correlation at the prescribed lag order.

Lag Order	Obs R-Squared Statistic	p-value
1st Order	2.23	0.1357
2nd Order	2.39	0.3033
3rd Order	2.95	0.3987
4th Order	3.81	0.4319
5th Order	4.96	0.4213
6th Order	6.21	0.3997

The next step in the model diagnostic stage involves the assessment of the stability of the model. This involves a close examination of the CUSUM test for coefficient stability and the CUSUM of squares test for stability in the variance of the model. The results of the two tests are shown below in figures 9 and 10. Figure 9 suggests that the model coefficients are stable, since the CUSUM plot lies within the 95 percent confidence interval. However, in figure 10 it appears that the model variance is not stable in certain years, since the plot of the CUSUM of squares statistic seems to lie outside the 95 percent confidence interval in some years.

Overall, based on the stability tests shown in figure 9 and 10, the model appears to be stable with a slight instability observed in the variance. The slight instability observed in the variance could potentially reduce the accuracy of the forecast; however, the SARDL model overall appears to be a model suitable enough to assess the impact oil prices will have on the unemployment rate forecast. As for the TARDL model, the model diagnostic tests resulted in the same outcomes as the SARDL model. It is important to note that the CUSUM of squares test showing variance instability for both the SARDL and the TARDL models might indicate that both ARDL models could potentially result in forecasts the might be inaccurate.

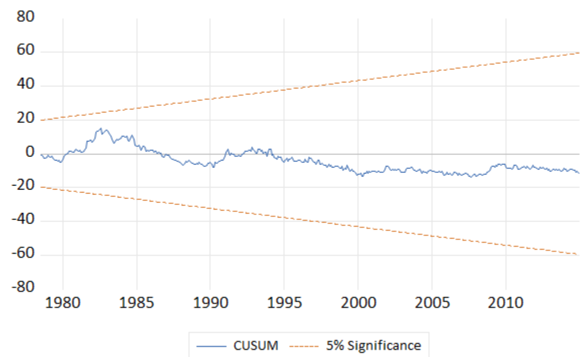


Figure 9: The CUSUM test, SARDL model.

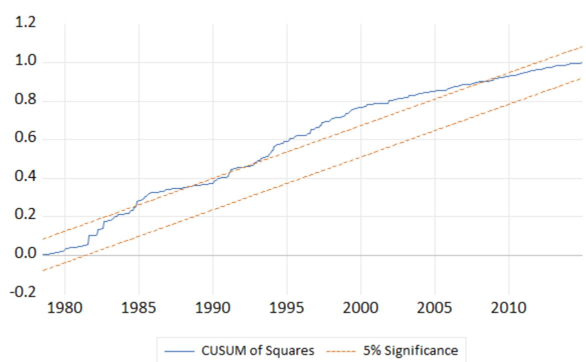


Figure 10: The CUSUM of squares test, SARDL model.

5.3 Forecasting Phase

The out of sample forecast involves a comparison between the seasonally diiferenced ARIMA model, the twice differenced ARIMA model, the seasonal random walk model, and the two ARDL models. First, a 1-step ahead forecast is carried out followed by a 3-steps ahead forecast with the estimation window increasing by one observation with each forecast.[18] The results of the forecasts are shown below in table 13. The forecasting performance is determined by which model achieves the lowest RMSE, MAE, and/or MAPE.

Table 13: Forecasting performance of the models.

Forecasting model/ Accuracy of Forecasts		(Horizon) 1-Step Ahead	(Horizon) 3-Steps Ahead
Seasonal Random Walk			
	RMSE	0.201039	0.307868
	MAE	0.154167	0.256522
	MAPE	2.404789	3.950015
SARIMA$(\|\|3,4,5\|\|,1,0)(0,1,1)_{12}$			
	RMSE	**0.173912**	0.25897
	MAE	**0.128696**	0.215705
	MAPE	**1.997603**	3.386948
TARIMA$(\|\|1,3,5\|\|,0,0)(0,1,1)_{12}$			
	RMSE	0.174937	**0.257181**
	MAE	0.131297	**0.211788**
	MAPE	2.048744	**3.348422**
SARDL $(\|\|1,6\|\|,\|\|4\|\|)$			
	RMSE	0.20678	0.329012
	MAE	0.158209	0.273214
	MAPE	2.460066	4.23071
TARDL $(\|\|3,4,12\|\|,\|\|3\|\|)$			
	RMSE	0.190023	0.296505
	MAE	0.14673	0.239447
	MAPE	2.276485	3.743961
No. of forecasts		48	45

[18]The forecast period for the 1-step ahead forecast is 2015M01 2018M12, and for 3-steps ahead it is 2015M03 2018M12. This is why the number of forecasts for the 1-step ahead is 48 and for the 3-steps ahead it is 45, as shown in table 13.

Table 14: The assessment of biasedness of the forecasts.

Model	Forecast Horizon	F Statistic	P value
Seasonal Random Walk	1 step	1.596	0.214
	3 step	4.980	0.011
TARIMA$(\|\|3,4,5\|\|,1,0)(0,1,1)_{12}$			
	1 step	0.15	0.862
	3 step	1.36	0.267
SARIMA$(\|\|1,3,5\|\|,0,0)(0,1,1)_{12}$			
	1 step	0.283	0.7549
	3 step	1.566	0.22
SARDL $(\|\|1,6\|\|,\|\|4\|\|)$			
	1 step	1.297	0.283
	3 step	5.236	0.009
TARDL $(\|\|3,4,12\|\|,\|\|3\|\|)$			
	1 step	0.813	0.45
	3 step	4.486	0.017

According to the results shown in table 13, based on the values of the root-mean-squared-error (RMSE), the mean-absolute-error (MAE), as well as the mean average percentage error (MAPE), the model that forecasts the unemployment rate the best during the 1-step ahead is the ARIMA$(||3,4,5||,1,0)(0,1,1)_{12}$ model (TARIMA). The model that forecasts the unemployment rate the best 3-steps ahead is the ARIMA$(||1,3,5||,0,0)(0,1,1)_{12}$ model (SARIMA). However, at this stage a judgment cannot be made about which ARIMA model is better until the statistical significance of the differences is assessed. Neither of the two ARDL models were better than the ARIMA models at forecasting the unemployment rate either 1-step ahead or 3-steps ahead. However, the best ARDL model in this paper was determined to be the ARDL$(||3,4,12||,||3||)$, or the TARDL model.

After the forecasting performance of the models is assessed, the biasedness of the forecasts is assessed through an F test. The results of the test are shown in table 14. Based on the results in table 14, the only two models that produce unbiased forecasts of the Canadian unemployment rate both 1 step ahead and 3 steps ahead are the SARIMA and TARIMA models. The seasonal random walk and both ARDL models produced unbiased forecasts for the 1 step ahead horizon but not the 3 step ahead horizon.

Table 15: Diebold-Mariano test for predictive accuracy.

	1 step ahead		3 step ahead	
Models Compared	**Squared**	**Absolute**	**Squared**	**Absolute**
1) Seasonal Random Walk 2) SARDL $(\|\|1,6\|\|,\|\|4\|\|)$	-0.582	-0.45	-0.994	-0.828
1) Seasonal Random Walk 2) TARDL $(\|\|3,4,12\|\|,\|\|3\|\|)$	1.082	0.626	0.529	0.813
1) SARIMA$(\|\|1,3,5\|\|,0,0)(0,1,1)_{12}$ 2) SARDL$(\|\|1,6\|\|,\|\|4\|\|)$	-1.761^{*}	-1.71^{**}	-2.743^{***}	-2.5^{***}
1) SARIMA$(\|\|1,3,5\|\|,0,0)(0,1,1)_{12}$ 2) TARDL$(\|\|3,4,12\|\|,\|\|3\|\|)$	-1.485	-1.483	-1.952^{*}	-1.389
1)TARIMA$(\|\|3,4,5\|\|,1,0)(0,1,1)_{12}$ 2) SARDL$(\|\|1,6\|\|,\|\|4\|\|)$	-1.812^{*}	-1.78^{*}	-2.616^{***}	-2.266^{***}
1)TARIMA$(\|\|3,4,5\|\|,1,0)(0,1,1)_{12}$ 2) TARDL$(\|\|3,4,12\|\|,\|\|3\|\|)$	-1.638	-1.723^{*}	-2.243^{*}	-1.292
1)SARDL$(\|\|1,6\|\|,\|\|4\|\|)$ 2) TARDL$(\|\|3,4,12\|\|,\|\|3\|\|)$	1.299	0.983	1.564	1.814^{*}

Note: In the first column, the first equation in each row is followed by the second equation.
$*$ significant at $p < 0.10$, $**$ significant at $p < 0.05$, $***$ significant at $p < 0.01$

To determine whether the observed differences in RMSE and MAE between the models being assessed is in fact significant, the Diebold-Mariano test of predictive accuracy is carried out. It is important to emphasisze that according to Diebold (2015), the Diebold-Mariano test is not supposed to be used to compare models, but rather, it is used to compare forecasts.

According to the Diebold-Mariano (DM) test results shown in table 15, the SARIMA model forecasts the unemployment rate better than the SARDL model both 1 step ahead and 3 steps ahead. The DM statistic is negative, which indicates that the SARIMA model is better, and the differences in RMSE and MAE between the SARIMA and SARDL models are significant. The TARIMA model also outperforms the SARDL model during the 1 step ahead and 3 steps ahead forecasts as shown by the results in table 15. The DM statistic here again is both negative and statistically significant, which indicates that the TARIMA model is better.

According to the Diebold-Mariano (DM) test results shown in table 15, the SARIMA model forecasts the unemployment rate better than the TARDL model both 1 step ahead and 3 steps ahead, as evidenced by the negative DM statistic. However, the differences in RMSE and MAE between the SARIMA and TARDL models are not significant in the 1 step ahead case. The difference in forecast accuracy is significant for the 3 steps ahead forecast at the 10 percent level, but only when the squared loss function is used. The TARIMA model also outperforms the TARDL model during the 1-step ahead and 3-steps ahead forecast horizons as shown by the results in table 15. The DM statistic here is again negative but unlike the results obtained with the SARIMA model, the DM statistic is statistically significant for the 1-step ahead forecast at the 10 percent level when the loss function is absolute but not when it is squared, and for the 3-steps ahead forecast at the 10 percent level when the loss function is squared but not when it is absolute, which suggests that the TARIMA model is indeed better at forecasting the unemployment rate than the TARDL model.

According to table 15, the seasonal random walk forecasts the unemployment rate better than the SARDL model, because the DM statistic is negative, but the results are not significant at any level. The seasonal random walk model also forecasts better than the SARDL model 3 steps ahead as evidenced by the negative DM statistic. However, the difference is not significant at any level. On the other hand, the TARDL forecasts better than the seasonal random walk during the 1 step ahead and 3 steps ahead forecasts as evidenced by the positive DM statistic. However, the differences are not statistically significant. Finally, the TARDL forecasts better than the SARDL model during the 1 step ahead and 3 steps ahead forecasts as evidenced by the positive DM statistic.[19] The results for the 1 step ahead forecasts are not significant. However, the DM statistic for the 3 steps ahead forecast is significant at the 10 percent level when the loss function absolute.

Overall, the DM test reveals that the ARIMA model which bests both ARDL models is the TARIMA. Based on what has been observed in table 15 as well as in table 13, the TARDL model is not only the best ARDL model in this paper, but it also fares a little bit better than the SARDL

[19] Since the TARDL model and SARDL model are not nested, the Diebold-Mariano test can be used to compare them.

model against both ARIMA models compared to the SARDL model. Both SARIMA and TARIMA models are better than the SARDL model. It is a little less clear that the SARIMA and TARIMA are a lot better than the TARDL. However, although the TARDL model appears to be a little better than the SARDL model, neither produced more accurate forecasts than the benchmark seasonal random walk model. The forecasting performance of the two ARIMA models cannot be compared with each other or the seasonal random walk using the DM test since they are all nested. In order to determine which ARIMA model forecasts the unemployment rate best, the Clark-West test is carried out.

Table 16: The Clark-West test of the predictive accuracy of two nested models. Regressions were conducted using Newey-West standard errors (HAC).

Models	Forecast Horizon	t Statistic	P value
1) seasonal random walk	1 step	3.078	0.004
2) SARIMA$(\|\|1,3,5\|\|,0,0)(0,1,1)_{12}$	3 step	3.396	0.001
1) Seasonal Random walk	1 step	2.906	0.006
2) TARIMA$(\|\|3,4,5\|\|,1,0)(0,1,1)_{12}$	3 step	3.341	0.002
1) SARIMA$(\|\|1,3,5\|\|,0,0)(0,1,1)_{12}$	1 step	1.156	0.253
2) TARIMA$(\|\|3,4,5\|\|,1,0)(0,1,1)_{12}$	3 step	0.443	0.660

As shown in table 16, both ARIMA models out perform the benchmark seasonal random walk for the 1-step ahead and 3 steps ahead forecasts. The t statistic in all four cases is greater than 1.645 and thus significant at the 10 percent level, which clearly indicates that the null hypothesis of equal predictive accuracy is rejected and the better models are indeed the ARIMA models. Between the two ARIMA models, it appears that both models predict the Canadian unemployment rate equally well. Therefore, any differences in RMSE, MAE, MAPE between the two ARIMA models could simply be due to the luck of the draw. It cannot be concluded that either one of the models is better than the other.

Figures 11, 12, and 13 shown below illustrate how the predicted values of the best model at each forecast horizon as determined by the RMSE, the MAE, and the MAPE compare to the actual values of the Canadian unemployment rate. For purposes of comparison, the one-step ahead forecast generated by the TARDL model is shown in figure 13. Overall, the forecasts of each model seem to be fairly close to the actual values.

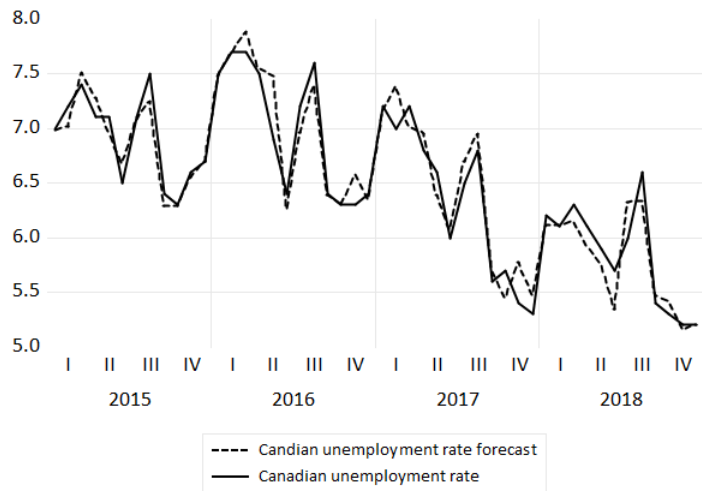


Figure 11: The TARIMA $(||3,4,5||,1,0)(0,1,1)_{12}$ 1 step ahead forecast of the Canadian unemployment rate.

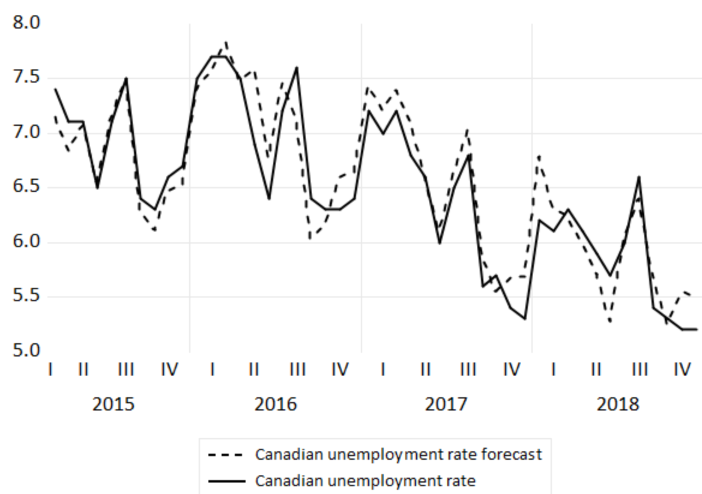


Figure 12: The SARIMA $(||1,3,5||,0,0)(0,1,1)_{12}$ 3 step ahead forecast of the Canadian unemployment rate.

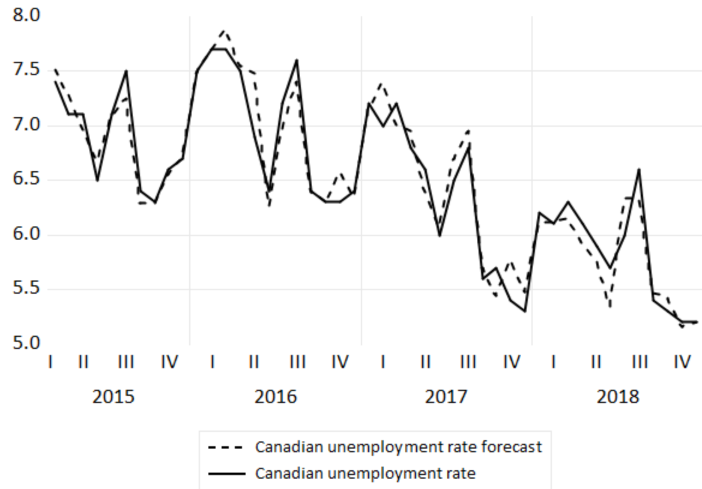


Figure 13: The TARDL($||3,4,12||,||3||$) 1 step ahead forecast of the Canadian unemployment rate.

6 Conclusion

For the most part, the models that most accurately forecast the Canadian unemployment rate are the two ARIMA models. Not only do they produce superior forecasts, but their forecasts are also unbiased. Both ARDL models produced unbiased forecasts of the Canadian unemployment rate for 1-step ahead, but not 3-steps ahead. Neither ARDL model produced forecasts that were in any way superior to the seasonal random walk. However, the best ARDL model in this paper appears to be the TARDL model.

The poor performance of both ARDL models overall raises some doubt about suitability of using the oil price as a leading indicator of the Canadian unemployment rate. However, the poor performance may just be attributed to the ARDL model itself. Perhaps additional leading indicators, or a much better specified model could potentially lead to the development of a more successful model that incorporates the oil price. During the ARDL model diagnostic stage, the CUSUM of squares test revealed variance instability. Perhaps future research that utilizes this ARDL model could augment the model to make the variance more stable. This could perhaps enhance the accuracy of the ARDL model forecasts especially for longer horizons beyond 1-step ahead.

The oil industry is a very significant part of the Canadian economy. It not only creates high paying jobs, but it is also an important source of revenue for the Canadian provinces that produce oil. The TARDL model in this paper can be used to forecast the Canadian unemployment rate 1-step ahead by including movements in oil prices as an exogenous variable. What this means is that if forecasters predict oil prices to be low, then it is reasonable to assume that in the short run at least, the unemployment rate will rise.

Other papers in the literature have used variables such as the Consumer Price Index (CPI) and the Industrial Producer Price Index (IPPI) successfully; however both of these variables could not be used to forecast the Canadian unemployment rate in this paper since during the model diagnostic stage, each of these variables led to misspecified models that could not be successfully corrected with further inclusion of these variables. Furthermore, the models explored in this paper do not have a built in mechanism to accommodate non-linearities that could exist in the unemployment rate time series. It has been shown in the literature that such non-linearities do exist in unemployment rate time series of other countries. So perhaps a TAR model that resembles the model used by Montgomery et al. (1998) might be used for this series to see if incorporating nonlinearities will lead to better forecasts.

7 References

Barnichon, R., C. J. Nekarda, J. Hatzius, S. J. Stehn, and B. Petrongolo (2012). The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market [with Comments and Discussion]. *Brookings Papers on Economic Activity*, 83–131.

Będowska-Sójka, B. (2015). Unemployment Rate Forecasts: Evidence from the Baltic States. *Eastern European Economics 53*(1), 57–67.

Beaulieu, J. and J. A. Miron (1993). Seasonal unit roots in aggregate U.S. data. *Journal of Econometrics 55*(1), 305–328.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society. Series B (Methodological) 37*(2), 149–192.

Claveria, O. (2019). Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations. *Journal for Labour Market Research 53*(1), 3.

Diebold, F. X. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. *Journal of Business & Economic Statistics 33*(1), 1–1.

Dritsakis, N. and P. Klazoglou (2018). Forecasting Unemployment Rates in USA using Box-Jenkins Methodology. *International Journal of Economics and Financial Issues 8*(1), 9–20.

Edlund, P.-O. and S. Karlsson (1993). Forecasting the Swedish unemployment rate VAR vs. transfer function modelling. *International Journal of Forecasting 9*(1), 61–76.

Enders, W. (2014). *Applied Econometric Times Series*. Wiley Series in Probability and Statistics. New York, United States: Wiley.

Früiiwirth-Schnatter, S. (1996). Recursive residuals and model diagnostics for normal and non-normal state space models. *Environmental and Ecological Statistics 3*(4), 291–309.

Funke, M. (1992). Time-series forecasting of the German unemployment rate. *Journal of Forecasting 11*(2), 111–125.

Hylleberg, S., R. F. Engle, C. W. J. Granger, and B. S. Yoo (1990). Seasonal integration and cointegration. *Journal of Econometrics 44*(1), 215–238.

Jalles, J. T. (2017). Forecasting performance of private sector's unemployment forecasts in advanced economies. *International Review of Applied Economics 31*(6), 707–733.

Jung, Y. C. and A. Das (2018). Asymmetric effects of the WTI crude oil price on unemployment rates: a comparative study of Canadian provinces and the United States. *OPEC Energy Review 42*(2), 123–150.

Khan Jaffur, Z. R., N.-U.-H. Sookia, P. Nunkoo Gonpot, and B. Seetanah (2017). Out-of-sample forecasting of the Canadian unemployment rates using univariate models. *Applied Economics Letters 24*(15), 1097–1101.

Kim, S. and H. Kim (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting 32*(3), 669–679.

Ledolter, J. and B. Abraham (1981). Parsimony and its importance in time series forecasting. *Technometrics 23*(4), 411–414.

Mahmudah, U. (2017). Predicting unemployment rates in Indonesia. *Economic Journal of Emerging Markets 9*(1), 20–28.

Montgomery, A. L., V. Zarnowitz, R. S. Tsay, and G. C. Tiao (1998). Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association 93*(442), 478–493.

Neusser, K. (2016). *Time Series Econometrics*. Springer Texts in Business and Economics. Cham: Springer International Publishing.

Ng, S. and P. Perron (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica 69*(6), 1519–1554.

Oil Sands Magazine (2019). *Oil Price Differentials Explained: Why Alberta crude sells at a deep discount*. (accessed February 21, 2019). https://www.oilsandsmagazine.com/market-insights/crude-oil-pricing-differentials-why-alberta-crude-sells-at-deep-discount-to-wti.

Perron, P. and S. Ng (1996). Useful Modifications to some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties. *Review of Economic Studies 63*(3), 435–463.

Stock, J. and M. Watson (2012). *Introduction to Econometrics*. Addison-Wesley series in economics. Upper Saddle River, United States: Pearson Education.

Stock, J. H. and M. W. Watson (2017). Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives 31*(2), 59–86.

Wilson, T. (1985). Lessons of the Recession. *Canadian Journal of Economics / Revue canadienne d'Economique 18*(4), 693–722.