Forecasting the Spread between the 'Real' and Official Unemployment Rates in Canada: Optimizing Forecast Performance of Multiplicative Seasonal ARIMA Models

By Sergei S. Shibaev
5739132

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

Supervisor: Louis-Philippe Morin

ECO 7997

Ottawa, Ontario
April 2011

Abstract

In this paper I look at the spread between the 'real' and official unemployment rates. The spread accounts for individuals that are marginally attached to the labour force (i.e. individuals discouraged by market conditions), and is therefore capable of capturing labour market impacts that follow contractionary periods in the business cycle. Analyzing the spread is useful to assess the true unemployment picture in Canada which may not be accurately portrayed by the official unemployment rate. I deliver a semi-automatic model-mining exercise that optimizes the predictive power of a model specification by focusing on minimizing out-of-sample forecast errors based on a group of forecast evaluation statistics. I use various penalty function criteria as objective measures to identify a portfolio of autoregressive integrated moving average models with a structural composition of seasonal and non-seasonal parts. I find that for the spread time series, the procedure advocates that, out of 576 estimated models, $ARIMA(1,0,1) \times (1,0,1)_{12}$ is not only the best fitted process, but also generates the top performing forecasts.

1 *Introduction*

The recent financial crisis, which began in the third quarter of 2008, has had an interesting impact on the Canadian labour market. Following the recovery of the economy, the official and supplementary unemployment rates (UR's) have all decreased from their peak levels during the recession. However, the gap between the two rates has remained at a constant and higher level for two consecutive years (the past two years), than that observed before the crisis. This impact on the labour market is not captured by the official UR, which is the standard indicator for assessing the well-being of the labour market. This forms a basis for exploring what can be expected or predicted in the dynamic of this gap between UR's (from now on referred to as the 'spread') for the short-term horizon.

The general structure and analysis of the paper follow that of Meyler, Geoff and Quinn (1998), who use multiplicative seasonal autoregressive integrated moving average models (ARIMA) to forecast inflation in Ireland and build a semi-automatic ARIMA modelling algorithm. This paper builds on their approach by adding forecast evaluation statistics that scale the forecast mean errors to a percentage of the realized values, which delivers a clearer evaluation of a forecasts precision. I also include an outlook for the spread 12 months into the future based on the predictions of the ARIMA process favoured by the exercise.

The methodology used in this paper, ARIMA time series modeling, is widely used to forecast variables of interest ranging from national economic accounts to prices. In relating current to previous values of a variable, univariate time series models have proven to form reliable short-term forecasts (Johnston and DiNardo 1984). ARIMA models have been commonly used in the literature to forecast the UR (e.g. Montgomery et al. (1998); Dobre and Alexandru (2008); and Funke (1992)). The UR has also been forecasted by variations of ARIMA, such as fractionally-integrated autoregressive moving average models (ARFIMA) (e.g. Kurita (2010)). The performance of ARIMA models can also serve as a benchmark for other approaches, as seen in Wilson and Perry (2004), where ARIMA models are used as reference points for spectral analysis models.

To one's surprise, there is little or no evidence in the literature that suggest a series similar to the spread has been forecasted using any particular approach for the Canadian context. Given the interesting post-crisis development in the Canadian labour market (mentioned earlier), modelling and forecasting the spread for the post-crisis period is not only of particular interest, but is a great entry point to introduce such a study into the literature. My paper provides a benchmark for modelling and forecasting the spread, and can be used to assess the relative forecast performance of other approaches.

The data series examined is the spread between the 'real'[1] and official UR's for Canada. The main difference between the official and the 'real' unemployment rates lies in their definition of an 'unemployed' individual. The 'real' rate refers to the official UR that also includes discouraged workers, waiting groups and involuntary part-timers. The official rate captures persons that were on temporary layoff with an expectation of recall, without work but had actively looked for work in the past four weeks or had a new job to start within four weeks (Statistics Canada, 2008). Given this definition, the official rate does not take into account all individuals that are affected during contractionary periods in the

[1] In the literature the term 'real' is interchangeably used with 'hidden' and 'underutilized.'

business cycle. In particular - the number of discouraged workers, waiting groups[2] and involuntary part-timers - persons that are marginally attached to the labour force. In particular, as concluded by Riddell (1999), who addressed the definition and measurement of unemployment, waiting groups in Canada display a strong labour force attachment. By including all marginally attached workers in the 'real' unemployment series, I account for all sub-categories of the marginal attachment population. This enables my analysis of the spread to relate back to these groups at the aggregate.

The aforementioned labour groups carry social costs (i.e. social assistance) and tend to increase significantly as labour market conditions worsen. Ernest B. Akyeampong (1992) traces the trends in discouraged workers following the early 1990's recession in Canada, while emphasising the importance of looking at the population of this group, which is not highly publicized. The observed number of discouraged workers had increased for two consecutive years following the early 1990's recession. Because the spread captures persons discouraged by market conditions, the preceding view contributes to my notion that the spread serves as a useful indicator of the true unemployment picture in Canada following the recent recession.

The objective of this paper is to deliver a semi-automatic model-mining exercise that optimizes the predictive power of a model specification by focusing on minimizing out-of-sample forecast errors based on a group of forecast evaluation statistics. I use various penalty function criteria as objective measures to identify a portfolio of ARIMA models with a structural composition of seasonal and non-seasonal parts. In optimizing the predictive

[2] These are persons that are available for work and are waiting for employment: waiting for recall to a former job, waiting for replies or waiting to start a new job (Riddell (1999)).

power of a model for the spread, I provide a basis for analyzing the future dynamic of unemployment in Canada.

The structure of this paper is as follows. In sections 2 and 3, I deliver an overview of multiplicative seasonal ARIMA processes and their structure, followed by an in-depth examination of the time-series in question – the spread. In section 4, I first present an objective approach to discriminating between the estimated 576 multiplicative seasonal ARIMA processes for the spread. I then carry forward a portfolio of top ranking processes and conduct various diagnostic checks. Lastly, I deliver a three-step-ahead forecasting exercise and compute various forecast evaluation statistics from the results. The final section of the paper, section 5, discusses what one can infer from the results pertaining to unemployment in Canada.

2 *Overview of ARIMA Models*

2.1 *Structure*

For seasonal processes, the polynomials of a general $ARIMA(p,d,q)$ process are factorized into two operators L and L^s (s is the seasonal span). A general seasonal autoregressive integrated moving average process with a constant, denoted by $ARIMA(p,d,q) \times (P,D,Q)_s$, is written as

$$\phi(L)\Phi(L^s)\nabla^d\nabla_s^D x_t = \alpha + \theta(L)\Theta(L^s)\varepsilon_t$$

and similarly, for an $I(0)$ process $ARIMA(p,0,q) \times (P,0,Q)_s$

$$\phi(L)\Phi(L^s)x_t = \alpha + \theta(L)\Theta(L^s)\varepsilon_t$$

where the lag operators are defined as

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

$$\Phi(L^s) = 1 - \Phi_s L^s - \Phi_{2s} L^{2s} - \cdots - \Phi_{Ps} L^{Ps}$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

$$\Theta(L^s) = 1 + \Theta_s L^s + \Theta_{2s} L^{2s} + \cdots + \Theta_{Qs} L^{Qs}$$

This representation is a multiplicative seasonal ARIMA. The lag operators could also be expressed as a single polynomial which would result in an additive seasonal ARIMA. Both the multiplicative and additive models attempt to capture seasonality. I use the multiplicative specification because it is able to capture more intricate seasonality dynamics in the series without adding more parameters (Enders, 2010).

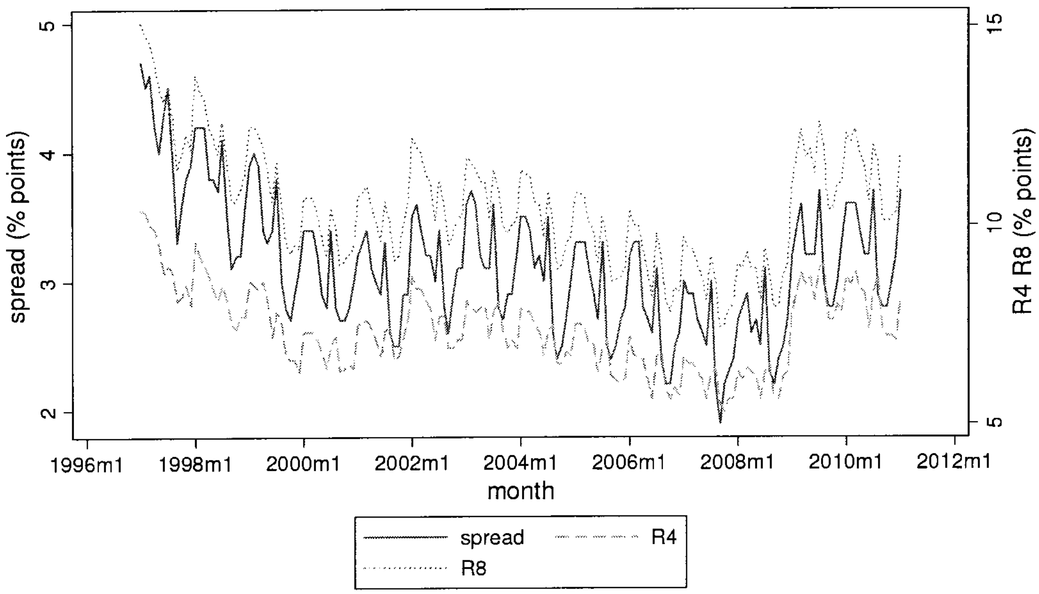
3 *Data*

3.1 *Data Collection and Examination*

This paper uses the spread in unemployment rates, x_t, constructed by taking the difference between the 'real' unemployment rate (R8) and the official unemployment rate (R4). The R8 is constructed by adding to the R4 those who are marginally attached to the labour force (i.e. discouraged workers, waiting groups, involuntary part-timers). The data is retrieved from Statistics Canada's monthly Labour Force Survey (LFS) which is the official source of unemployment rate statistics in Canada. Each series is reported monthly, unadjusted for seasonality and only available for the period from January(m1) 1997 to January(m1) 2011. The resulting series consists of 169 observations.

$$spread: \quad x_t = \mathrm{R8}_t - \mathrm{R4}_t \quad , where\ t\ is\ the\ respective\ month\ of\ a\ given\ year$$

Figure 1
Spread, R4 and R8



It is evident from Figure 1 that all three series exhibit strong annual (12-month) seasonal patterns. The spread consistently peaks in the first three months of a year and then again every July. From the summary statistics reported in Table 1, it is evident that the mean of the spread have been slightly lower in the 2003m1-2011m1 period, while the standard deviation has been higher. This suggests that although R4 could be better reflecting true labour market conditions, the spread has become more volatile. However, from the plot of the series, there is no evidence of any outliers or structural breaks over the observed time period.

Table 1
Summary Statistics for the Spread Series

Sample Period	Mean	Standard Deviation	Observations
1997m1 - 2002m12	3.38	0.43	72
2003m1 - 2011m1	2.94	0.54	97
1997m1 - 2011m1	3.13	0.53	169

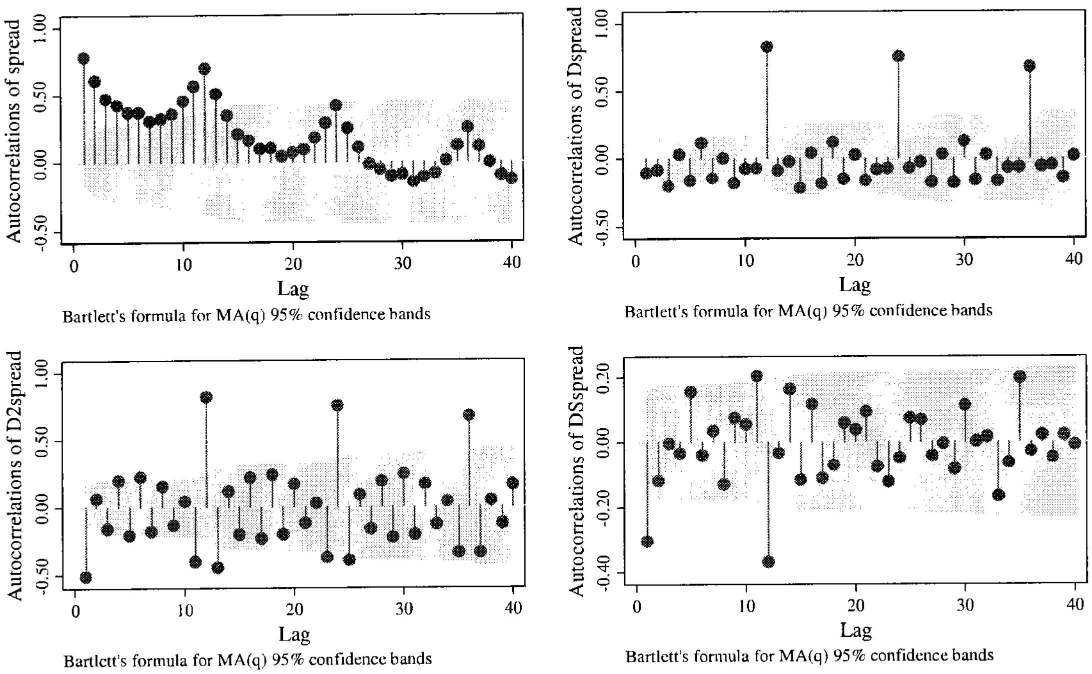
3.2 *Testing for Stationarity of Time Series*

Before moving to the model identification and estimation stages, I must check whether the series is stationary.[3] The series needs to be stationary in order to forecast; otherwise the estimates would not be reliable (Johnston and DiNardo 1984). I have already shown in Figure 1 that the spread series appears to be stationary. I can also look to the autocorrelogram for a given series. The plotted sample autocorrelations should indicate whether or not there is a nonstationary problem. In addition to the original spread series, I examine its first, second and seasonal difference transformations (see Figure 2). The maximum lag length considered for the sample autocorrelations is 40, which is just below the suggested maximum of $n/4$.[4] The sample autocorrelations of a stationary series should die out quickly from lag zero.

[3] From here, I will use the 1997m1-2010m7 sample of the spread series. The removed six observations (2010m8-2011m1) will allow me to evaluate the forecasted series using various forecast evaluation statistics (which are computed based on these six observations).
[4] When looking at the sample autocorrelations of a series, the rule of thumb is that lags exceeding twice the seasonal span of a series (in this case 24) should be considered with caution (Meyler et al. (1998)).

Figure 2

Sample Autocorrelations



The autocorrelations for the spread die out slowly in a strong seasonal pattern. The first and second differences of the spread, D.spread and D2.spread respectively, appear to be stationary but have significant and gradually decaying autocorrelations at lags 12, 24 and 36, exhibiting the same distinctive seasonality as the original series. The first-differenced seasonal difference of the spread, DS.spread, does not display any seasonality, but suggests over-differencing due to increased volatility in the autocorrelations (therefore it is not considered further).[5]

[5] When moving from one difference order to the next (higher order), then the variable is said to be over-differenced once the sample variance increases (Meyler et al. (1998) p.10).

Now that I have a better understanding of the behaviour of the spread and its transformations from their respective plotted sample autocorrelations, I carry the three series forward and perform formal unit root tests to determine stationarity.

3.2.1 *Test for Unit Root*

I formally test for the presence of unit roots to determine whether the spread, D.spread or D2.spread series are stationary. The Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests, test for the presence of a unit root without treating seasonality. To make sure that I don't falsely reject unit root, I also perform the Augmented Dickey-Fuller test that treats seasonality (SADF), when the seasonal pattern is purely deterministic (Enders 2010).

The ADF and PP report p-values of zero for the D.spread and D2.spread series, for which I reject unit root. The SADF reports test statistics of -2.96, -18.35 and -26.87 for the spread, D.spread and D2.spread respectively, again rejecting the presence of a unit root. I reject unit root for the spread series and both of the transformations, D.spread and D2.spread, therefore there is no need to de-trend the original spread series.

Table 2
Unit Root Tests

	Test Statistic	MacKinnon approximate p-value	
Time Series	*Augmented Dickey-Fuller with Seasonality*	*Augmented Dickey-Fuller*	*Phillips-Perron*
Spread	-2.96*	0.0002	0.0001
D.spread	-18.35**	0.0000	0.0000
D2.spread	-26.87**	0.0000	0.0000

*significant at 5% level
**significant at 1% level

3.2.2 Sample Variance

Even though I have now formally established that the original spread series is stationary, I take a look to the sample variances of the series to understand why the formal tests were stronger for transformations of the spread. In general, the sample variance of a series decreases as the difference level increases, until the correct difference order is reached. If the variance increases, when moving to a higher difference order, it is evidence of over-differencing (Meyler et al. 1998). The sample variance decreases as I move from the spread to D.spread and increases as I move from D.spread to D2.spread. If I adhered to this approach only, then the D.spread series would be carried forward since it displays the lowest sample variance.

Table 3
Sample Variances

Time Series	Sample Variance
spread	0.228
D.spread	0.107
D2.spread	0.236

4 *Multiplicative Seasonal ARIMA Modelling*

I know deliver an algorithm for specifying a multiplicative seasonal ARIMA process with the goal of minimizing out-of-sample forecast errors. I will start by using a variety of objective penalty function criteria for model identification and estimation, then perform a series of diagnostics checks and finally perform and evaluate the predicted values generated by each process.

4.1 *Model Specification and Estimation*

4.1.1 *Comparing Information Criterions*

Information criterions are objective measures that discriminate against model specifications with high residual sum of squares and number of parameters. I consider three information criterions, the Akaike Information Criterion (AIC), the Hannan-Quinn Information Criterion (HQIC) and the Bayesian Information Criterion (AIC). The three have a similar structure which is captured by the general form

$$Information\ Criterion = LL + f(k, n)$$

where LL is the log of the maximum likelihood estimation of the residual sum of squares, $f(k, n)$ is the penalty term that penalizes models for including more parameters k and n is the number of observations. It is clear that for $n > 7$ the BIC penalizes a model more than the AIC for including more parameters, meaning I can expect the AIC to favour higher order specifications.

4.1.2 *Objective Model Identification*

I set the maximum order of an ARIMA process at $ARIMA(11,0,11) \times (1,0,1)$, which is based on the rule of thumb given by

$$maximum\ order\ ARIMA = ARIMA(seasonal\ span - 1, 0, seasonal\ span - 1) \times (1,0,1)$$

(Meyler et al. 1998)

Generating values for the information criterions is computationally expensive. Having set this limit requires that I estimate 576 different ARIMA processes for which 576 AIC, BIC and HQIC values are reported:

Table 4

Composition of the maximum order ARIMA processes considered

Non-Seasonal Component	Seasonal Component	Number of Models
from (0,0,0) *to* (11,0,11)	(0,0,0)	144
from (0,0,0) *to* (11,0,11)	(1,0,0)	144
from (0,0,0) *to* (11,0,11)	(0,0,1)	144
from (0,0,0) *to* (11,0,11)	(1,0,1)	144

Table 5 reports the top 15 models specifications according to the information criterions.

Table 5
Model Ranking by Information Criterion

Rank	BIC		AIC		HQIC	
1	(1,0,1)x(1,0,1)*	-1.06	(1,0,1)x(1,0,1)	-1.155	(1,0,1)x(1,0,1)	-1.116
2	(2,0,0)x(1,0,1)	-1.051	(2,0,0)x(1,0,1)	-1.146	(2,0,0)x(1,0,1)	-1.108
3	(1,0,2)x(1,0,1)	-1.029	(1,0,2)x(1,0,1)	-1.143	(1,0,2)x(1,0,1)	-1.097
4	(3,0,0)x(1,0,1)	-1.025	(4,0,0)x(1,0,1)	-1.141	(3,0,0)x(1,0,1)	-1.093
5	(2,0,1)x(1,0,1)	-1.018	(1,0,5)x(1,0,1)	-1.14	(2,0,1)x(1,0,1)	-1.086
6	(4,0,0)x(1,0,1)	-1.008	(3,0,0)x(1,0,1)	-1.139	(1,0,3)x(1,0,1)	-1.085
7	(1,0,3)x(1,0,1)	-1.006	(1,0,3)x(1,0,1)	-1.139	(1,0,4)x(1,0,1)	-1.075
8	(1,0,4)x(1,0,1)	-0.985	(8,0,0)x(1,0,1)	-1.138	(5,0,0)x(1,0,1)	-1.071
9	(5,0,0)x(1,0,1)	-0.98	(1,0,4)x(1,0,1)	-1.136	(1,0,5)x(1,0,1)	-1.07
10	(1,0,0)x(1,0,1)	-0.979	(5,0,1)x(1,0,1)	-1.134	(5,0,1)x(1,0,1)	-1.065
11	(2,0,3)x(1,0,1)*	-0.97	(6,0,0)x(1,0,1)	-1.134	(6,0,0)x(1,0,1)	-1.064
12	(4,0,1)x(1,0,1)*	-0.97	(2,0,1)x(1,0,1)	-1.132	(4,0,1)x(1,0,1)	-1.06
13	(1,0,5)x(1,0,1)*	-0.969	(5,0,0)x(1,0,1)	-1.132	(2,0,3)x(1,0,1)	-1.06
14	(3,0,2)x(1,0,1)*	-0.965	(8,0,1)x(1,0,1)	-1.129	(3,0,2)x(1,0,1)	-1.055
15	(5,0,1)x(1,0,1)	-0.964	(2,0,5)x(1,0,1)	-1.128	(8,0,0)x(1,0,1)	-1.054

* Top 5 models that do not exhibit serial correlation in the residuals

Due to the attractive properties of the BIC discussed earlier, I focus on the BIC ranking only. Since the focus of the paper is on forecast performance, I need to construct a model of portfolios to carry forward. At first I apply the posterior odds ratio $\mathfrak{R}$ of the ARIMA specification, introduced by Poskitt and Tremayne (1987), which in this case is

$$\mathfrak{R} = exp\left[-\frac{163}{2}\{-1.06 - BIC(p,0,q)\}\right]$$

Where $-1.06 = BIC(p^*,0,q^*)$ is the BIC of the superior ranking model. Poskitt and Tremayne (1987) suggest that any process for which $1 < \mathfrak{R} < \sqrt{10}$, is a close competitor to the top ranking model. This means that for any model that reports a BIC within $2\log(\sqrt{10})/n = 0.00613$ of $ARIMA(1,0,1) \times (1,0,1)_{12}$ should be included in the portfolio. However, none of the estimated models fall within this range. I thus take a more straightforward approach and construct a portfolio consisting of the top 5 ranking models by the BIC to the diagnostics and forecasting stages. The models included in the portfolio are

Table 6
Portfolio of ARIMA Processes

Rank/ Model#	ARIMA Process	Specification	BIC
1	$ARIMA(1,0,1) \times (1,0,1)_{12}$	$(1-\phi_1\mathrm{L})\Phi(L^{12})x_t = \alpha + (1+\theta_1\mathrm{L})\Theta(L^{12})\varepsilon_t$	-1.06
2	$ARIMA(2,0,0) \times (1,0,1)_{12}$	$(1-\phi_1\mathrm{L}-\phi_2L^2)\Phi(L^{12})x_t = \alpha + \Theta(L^{12})\varepsilon_t$	-1.051
3	$ARIMA(1,0,2) \times (1,0,1)_{12}$	$(1-\phi_1\mathrm{L})\Phi(L^{12})x_t = \alpha + (1+\theta_1\mathrm{L}+\theta_2L^2)\Theta(L^{12})\varepsilon_t$	-1.029
4	$ARIMA(3,0,0) \times (1,0,1)_{12}$	$(1-\phi_1\mathrm{L}-\phi_2L^2-\phi_3L^3)\Phi(L^{12})x_t = \alpha + \Theta(L^{12})\varepsilon_t$	-1.025
5	$ARIMA(2,0,1) \times (1,0,1)_{12}$	$(1-\phi_1\mathrm{L}-\phi_2L^2)\Phi(L^{12})x_t = \alpha + (1+\theta_1\mathrm{L})\Theta(L^{12})\varepsilon_t$	-1.018

Note: $\Phi(L^{12}) = 1 - \Phi_{12}\mathrm{L}^{12}$ and $\Theta(L^{12}) = 1 + \Theta_{12}\mathrm{L}^{12}$

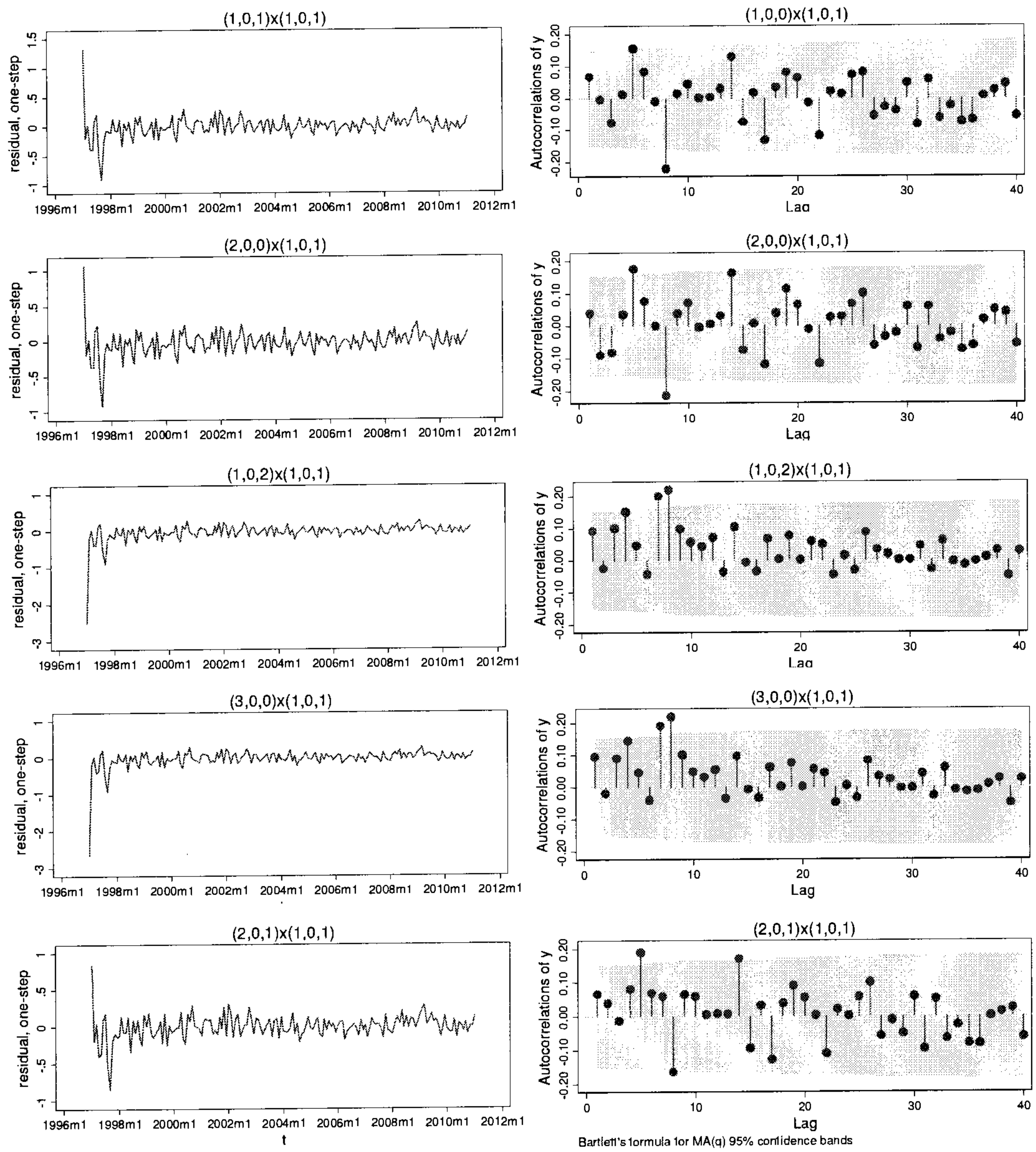
4.2 *Model Diagnostics*

4.2.1 *Autocorrelation in residuals*

Before moving to the forecasting stage, I perform diagnostic checks for the top ranking models. I start by plotting the residuals and the autocorrelogram of the residuals for each process.

Figure 3

Plotted Residuals and Autocorrelograms of the Residuals



If the models are correctly specified, then the behaviour of the residuals should be approximately white-noise (Johnston and DiNardo 1984). That is, if the residuals are white noise, then the corresponding autocorrelations should die out immediately after the first lag. Judging by the plotted residuals in Figure 3, the residuals for all five processes appear to be approximately white noise. When I look to the autocorrelation functions of each process; $ARIMA(1,0,1) \times (1,0,1)_{12}$ has only one significant autocorrelation at lag 8, $ARIMA(2,0,0) \times (1,0,1)_{12}$ has two significant autocorrelations at lags 5 and 8, $ARIMA(1,0,2) \times (1,0,1)_{12}$ and $ARIMA(3,0,0) \times (1,0,1)_{12}$have two significant autocorrelations at lags 7 and 8, and $ARIMA(2,0,1) \times (1,0,1)_{12}$ has one significant autocorrelation at lag 5.

Without formal testing, the only thing that I can conclude is that the residuals of all processes are approximately white noise, and that the $ARIMA(1,0,1) \times (1,0,1)_{12}$ and $ARIMA(2,0,1) \times (1,0,1)_{12}$ processes have the fewest significant autocorrelations (only for one lag). I now formally test for the presence of serial correlation in the residuals.

4.2.2 *Ljung-Box Test*

The Ljung-Box test checks for this problem for a fitted ARIMA process. The test can also be interpreted as testing for the overall randomness of the data for a given number of lags k.

$ARIMA(1,0,1) \times (1,0,1)_{12}$ is the only model specification for which I fail to reject the null since $p(Q(k) > \chi^2(k)) \geq 0.05$ for $k = 5, 10, 20, 30, 40$. For the other four models, the Ljung-Box test stipulates serially correlated residuals at lag 10 for $ARIMA(2,0,0) \times (1,0,1)_{12}$ and $ARIMA(3,0,0) \times (1,0,1)_{12}$ and lags 10 and 20 for $ARIMA(1,0,2) \times (2,0,1)_{12}$ (see appendix).

The test was performed for all subsequent models ranked by the BIC until four more models where found for which I failed to reject the null at lags $k = 5, 10, 20, 30, 40$. These models are ranked 11, 12, 13, and 14, and are marked with a * in Table 5. However, these models are not carried forward to the forecasting stage.

In the preceding section, 4.2.1, I plotted the residuals and the autocorrelogram of the residuals for the top five processes (by BIC). Visually examining the plots, I concluded that the residuals for all five processes appear to be approximately white noise, when in fact; judging by a formal test, the Ljung-Box test, from the five processes only the residuals for $ARIMA(1,0,1) \times (1,0,1)_{12}$ behave as a white noise process.

4.2.3 *Process Stationarity*

Noting the problem with the residuals for Models 2, 3, 4 and 5, indicated by the Ljung-Box test, I take Model 1, $ARIMA(1,0,1) \times (1,0,1)_{12}$ and test it for the unit root problem. I have to assert that the autoregressive polynomial operator has all the roots outside the unit circle so that the process is stationary (Hamilton 1994).

Model 1 $\quad ARIMA(1,0,1) \times (1,0,1)_{12}$

$$(1-\phi_1\mathrm{L})(1-\Phi_{12}\mathrm{L}^{12})x_t = \alpha + (1+\theta_1\mathrm{L})(1+\Theta_{12}\mathrm{L}^{12})\varepsilon_t$$

$$x_t(1-\phi_1\mathrm{L}-\Phi_{12}\mathrm{L}^{12}+\phi_1\Phi_{12}\mathrm{L}^{13}) = \alpha + \varepsilon_t(1+\theta_1\mathrm{L}+\Theta_{12}\mathrm{L}^{12}+\theta_1\Theta_{12}\mathrm{L}^{13}),$$

$$A(L) = 1-\phi_1\mathrm{L}-\Phi_{12}\mathrm{L}^{12}+\phi_1\Phi_{12}\mathrm{L}^{13},$$

$$B(L) = 1+\theta_1\mathrm{L}+\Theta_{12}\mathrm{L}^{12}+\theta_1\Theta_{12}\mathrm{L}^{13},$$

$$A(L)x_t = \alpha + B(L)\varepsilon_t$$

For the estimated model (see appendix)

$$x_t(1-0.9877361\mathrm{L}-0.9974887\mathrm{L}^{12}+0.9852556\mathrm{L}^{13})$$
$$= 3.990104 + \varepsilon_t(1-0.3848386\mathrm{L}-0.8103472\mathrm{L}^{12}+0.3113853\mathrm{L}^{13})$$

$Let\ L = \lambda^{-1}$

$$\lambda^{13} - \phi_1\lambda^{12} - \Phi_{12}\lambda + \phi_1\Phi_{12} = 0$$

$$\lambda^{13} - 0.9877361\lambda^{12} - 0.9974887\lambda + 0.9852556 = 0$$

Since all roots of $A(L)$ (see appendix) lie outside the unit circle (i.e. $\forall|\lambda| < 1$), I know that $A(L)$ is invertible, hence the representation

$$A(L)x_t = \alpha + B(L)\varepsilon_t$$

can be written as

$$x_t = \frac{\alpha}{(1-\phi_1)(1-\Phi_{12})} + \frac{(1+\theta_1 \mathrm{L})(1+\Theta_{12}\mathrm{L}^{12})}{(1-\phi_1 \mathrm{L})(1-\Phi_{12}\mathrm{L}^{12})}\varepsilon_t$$

which is the ARIMA(1,0,1)x$(1,0,1)_{12}$ process for the series $\{x_t\}$.

I have confirmed that all of the roots of the autoregressive polynomial operator lie outside the unit circle (i.e. $A(L)$ is invertible); however, some of the roots are not statistically different from 1. For my purpose, this is not a problem because I want to know the process is stationary in order to be able to forecast. Based on the formal test for unit root and the derivation of the polynomial roots I conclude that there is no unit root problem for $ARIMA(1,0,1) \times (1,0,1)_{12}$. The process is stationary and I can now move onto the forecasting stage of the analysis.

4.3 *Forecasting and Forecast Evaluation*

Since my primary focus is on optimizing forecast performance, I include models 1 to 5 in the forecasting exercise. I present a forecasting and forecast evaluation procedure that will complete the semi-automatic ARIMA forecasting algorithm.

4.3.1 *Three-Step-Ahead Forecast*

As discussed earlier, prior to the model identification stage, I stepped back six months in the available sample period in order to be able to conduct an in-sample forecast analysis. As such, I can perform six one-step-ahead forecasts, five two-step-ahead forecasts or four three-step-ahead forecasts. I deliver the forecasting exercise by conducting four three-step-ahead forecasts ranging from 2010m8-m10 to 2010m11-2011m1. Table 7 reports the forecast results.

Table 7

Three Step-ahead Predicted Values for the Spread

	Period 1 2010m8-m10	Period 2 2010m9-m11	Period 3 2010m10-m12	Period 4 2010m11-2011m1
STEP		$(1,0,1) \times (1,0,1)_{12}$		
1	2.956405	2.732202	2.888186	2.999738
2	2.737495	2.887371	3.001077	3.182021
3	2.889181	2.998715	3.182258	3.556467
STEP		$(2,0,0) \times (1,0,1)_{12}$		
1	2.964855	2.728242	2.882604	3.003402
2	2.727932	2.88152	3.003794	3.171308
3	2.884094	3.005712	3.168065	3.557152
STEP		$(1,0,2) \times (1,0,1)_{12}$		
1	2.967067	2.730949	2.889343	3.004291
2	2.731105	2.88998	3.001489	3.181063
3	2.896136	3.000539	3.182919	3.557972
STEP		$(3,0,0) \times (1,0,1)_{12}$		
1	2.953071	2.738776	2.891212	2.998204
2	2.731302	2.891677	2.999789	3.181428
3	2.891578	2.996986	3.192053	3.552583
STEP		$(2,0,1) \times (1,0,1)_{12}$		
1	2.960931	2.713825	2.885441	2.99624
2	2.73818	2.862589	2.995644	3.181393
3	2.896351	2.98739	3.185571	3.551741

Month	Observed value
2010m8	2.9
2010m9	2.8
2010m10	2.8
2010m11	3
2010m12	3.2
2011m1	3.7

I first graphically examine how the three-step-ahead predicted values perform for each process (see Figures 4 to 7). Visually, it is evident that the forecasts of all five processes behave very similarly. However, the analysis that follows will show that some of the processes clearly out-perform others.

Figure 4

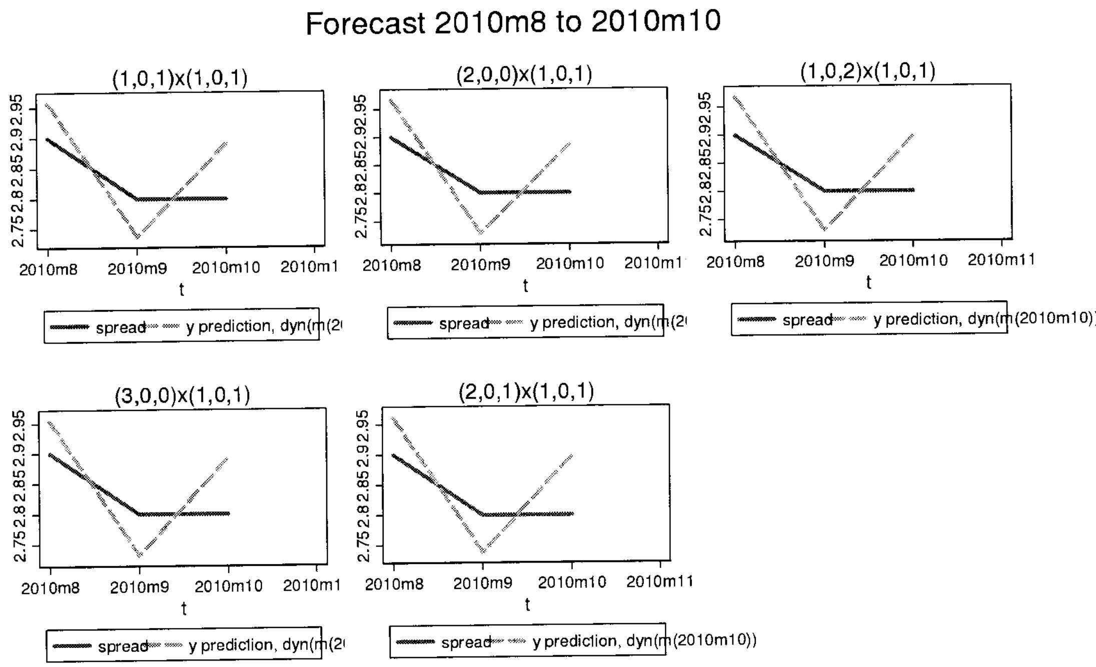


Figure 5

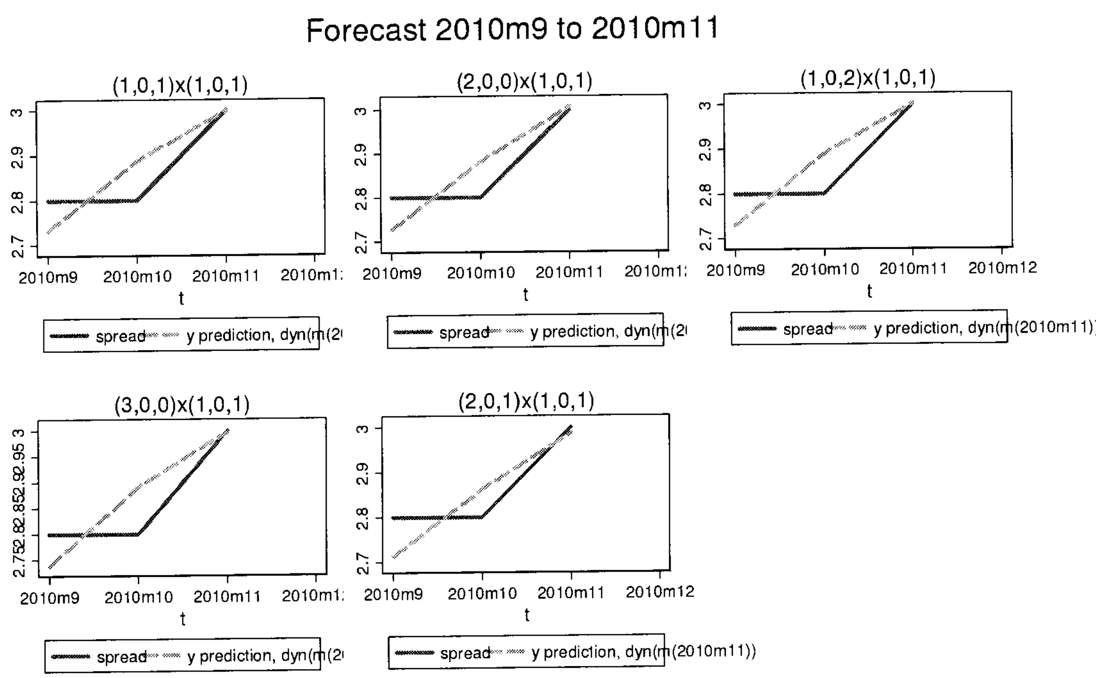


Figure 6

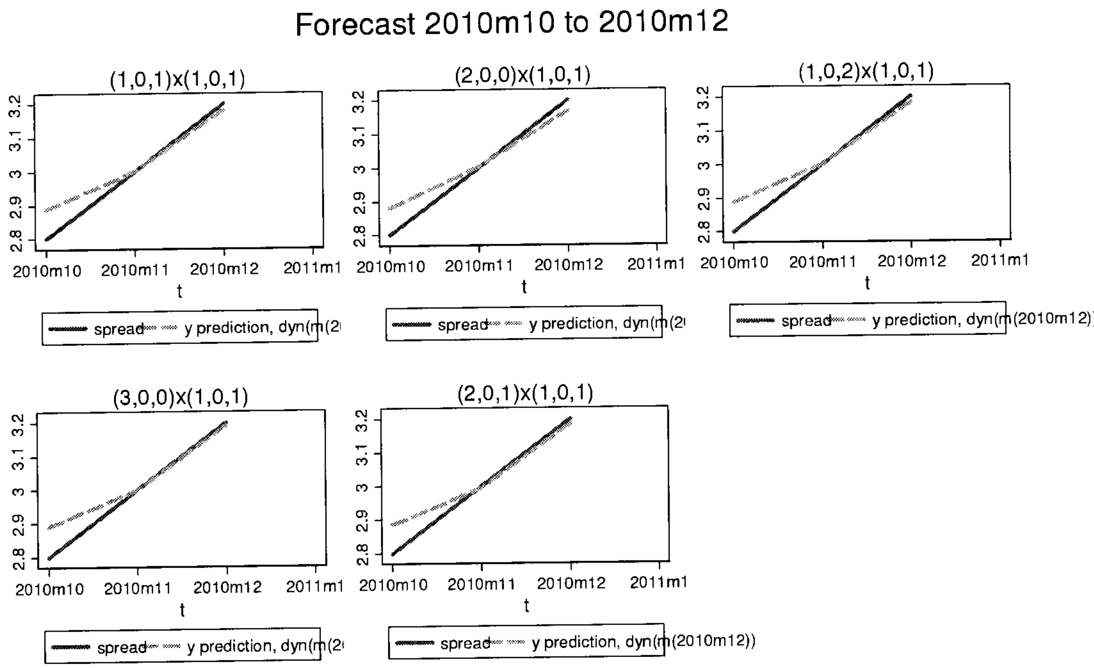
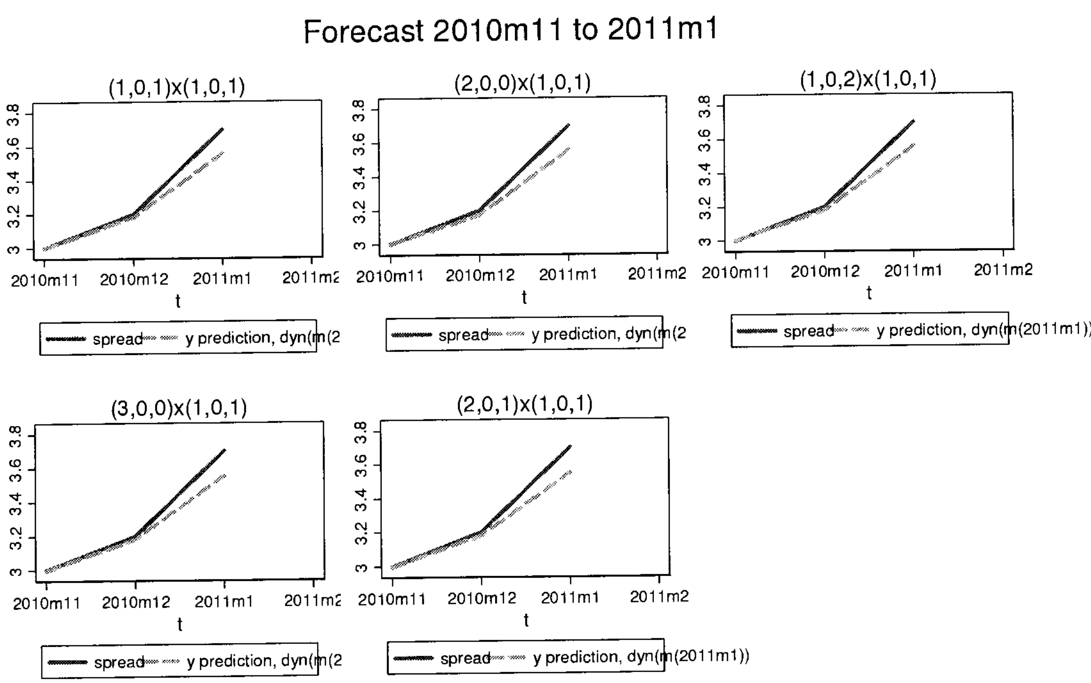


Figure 7



4.3.2 *Forecast Evaluation Statistics*

Following the preceding analysis, I now use various evaluation statistics to definitively discriminate among the five specified processes to determine which one process should be used for forecasting the spread series.

The forecast evaluation statistics examined are the Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE) and Theil's U-Statistic (TU).

Each statistic provides insight into the behaviour of the predictions each process generates. I will discuss what each statistic conveys using examples from the reported forecast evaluation statistics in Table 8 and then put forth the dominant process.

The ME and MPE are difficult measures to rely on because they can recommend a process with a ME or MPE of zero, when in fact that measure is capturing offsetting positive and negative errors (Cook 2006). However, the fact that the errors alternate in signs is a good occurrence because I know the process is not consistently overshooting or undershooting, this is referred to as forecasting bias (Meyler et al. 1998). If the process consistently forecasts too high, the ME and MAE or MPE and MAPE would be similar. Therefore, I use the ME to discriminate between processes not based on the value of the statistic, but on the presence of forecasting bias. From Figures 4 to 7, it is clear that none of the processes constantly forecast too high or low, meaning there is no forecasting bias present in any of them.

Both the MAE and the RMSE solve the inherent problem in the ME, but they fail to account for the scale of the series (Cook 2006). Fortunately, this problem is not relevant for my analysis since I am examining only one series. I deem the MAE and RMSE appropriate statistics to discriminate between processes based on their values – favouring the process with the lowest values for these statistics. The MAE favours $ARIMA(1,0,1) \times (1,0,1)_{12}$ for which $MAE = 0.0528$. The RMSE favours $ARIMA(1,0,1) \times (1,0,1)_{12}$ and $ARIMA(2,0,1) \times (1,0,1)_{12}$ equally, their RMSE's being 0.0673.

The MAPE provides useful information on the forecast by scaling the average error to a percentage of the realized value. The MAPE does not possess the shortcomings of the MPE and joins the MAE and RMSE in the analysis. The MAPE favours $ARIMA(2,0,1) \times (1,0,1)_{12}$ with a MAPE of 1.20. For all five processes the MAPE is between 1.20 and 1.85. This is an impressive result - the predictions of these processes on the three month horizon, on average, yield forecast percentage errors averaging between 1.2% and 1.85%. The second lowest MAPE is reported for $ARIMA(1,0,1) \times (1,0,1)_{12}$ with a value of 1.76.

The final statistic considered is TU. The TU is lowest for the most accurate forecasts and is bounded between 0 and 1 (Cook 2006). As did the RMSE, the TU favours $ARIMA(1,0,1) \times (1,0,1)_{12}$ and $ARIMA(2,0,1) \times (1,0,1)_{12}$ equally, for which the $TU = 0.0111$. All five processes display TU's between 0.0111 and 0.0125. Again, this suggests that the forecasting precision of these models is fairly high.

If I allowed for six decimal places in the reported forecast evaluation statistics, then the RMSE and TU would strictly favour $ARIMA(1,0,1) \times (1,0,1)_{12}$. However, since the RMSE and TU values are so similar for $ARIMA(1,0,1) \times (1,0,1)_{12}$ and $ARIMA(2,0,1) \times (1,0,1)_{12}$, it is fair to say that the two statistics favours both processes equally.

Table 8

Reported Forecast Evaluation Statistics

	ME	MAE	RMSE	TU	MPE	MAPE
STEP			$(1,0,1) \times (1,0,1)_{12}$			
1	-0.0191	0.0532	0.0624	0.0108	-0.6661	1.8811
2	-0.002	0.0422	0.0545	0.0092	-0.0979	1.4876
3	0.0183	0.0629	0.085	0.0134	0.3229	1.9154
Avg. 1-3	**-0.0009***	**0.0528***	**0.0673***	**0.0111***	**-0.1470**	**1.7614**
STEP			$(2,0,0) \times (1,0,1)_{12}$			
1	-0.0198	0.0557	0.0636	0.011	-0.6843	1.9657
2	0.0039	0.0465	0.0563	0.0095	0.1169	1.6271
3	0.0212	0.0661	0.0845	0.0133	0.4162	1.4331
Avg. 1-3	**0.0018**	**0.0561**	**0.0681**	**0.0113**	**-0.0504**	**1.6753**
STEP			$(1,0,2) \times (1,0,1)_{12}$			
1	-0.0229	0.0574	0.0657	0.0114	-0.7951	2.0282
2	-0.0009	0.0448	0.0575	0.097	-0.057	1.5789
3	0.0156	0.0639	0.0862	0.0135	0.2302	1.9559
Avg. 1-3	**-0.0027**	**0.0554**	**0.0698**	**0.0115**	**-0.2073**	**1.8543**
STEP			$(3,0,0) \times (1,0,1)_{12}$			
1	-0.0203	0.0518	0.061	0.0141	-0.7103	1.8335
2	-0.001	0.0448	0.058	0.098	-0.063	1.5788
3	0.0167	0.0625	0.0869	0.0137	0.2656	1.9009
Avg. 1-3	**-0.0016**	**0.053**	**0.0686**	**0.0125**	**-0.1692**	**1.7711**
STEP			$(2,0,1) \times (1,0,1)_{12}$			
1	-0.0141	0.0591	0.0679	0.0118	-0.4874	1.5846
2	0.0055	0.0368	0.045	0.0076	0.189	0.7626
3	0.0197	0.0679	0.0889	0.014	0.3593	1.254
Avg. 1-3	**0.0037**	**0.0546**	**0.0673***	**0.0111***	**0.0203***	**1.2004***

*Lowest statistic among the five processes

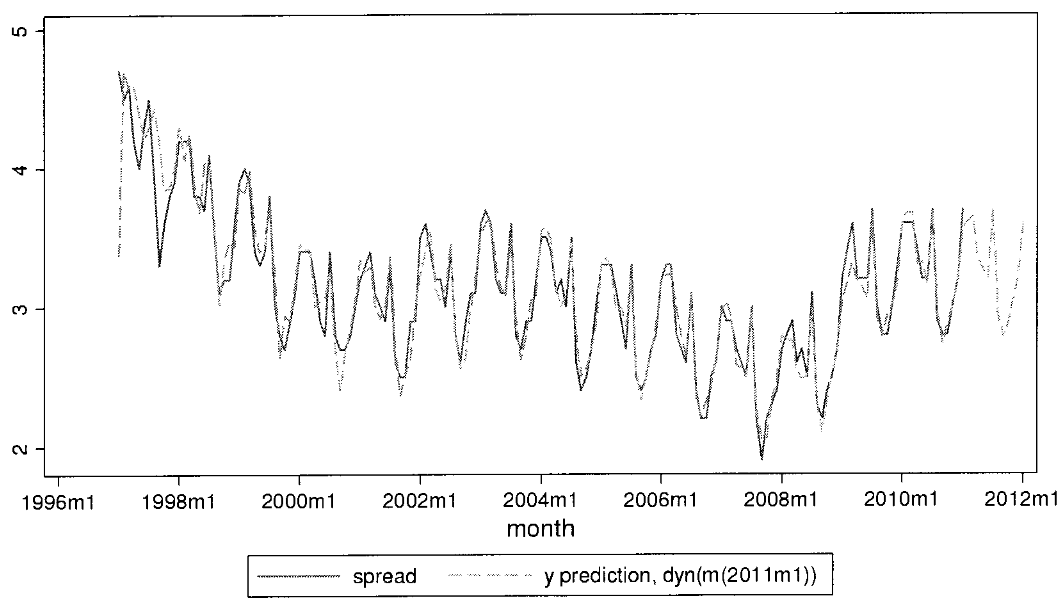
The forecast evaluation statistics favour $ARIMA(1,0,1) \times (1,0,1)_{12}$ and $ARIMA(2,0,1) \times (1,0,1)_{12}$. However, due to the unit root problem found for $ARIMA(2,0,1) \times (1,0,1)_{12}$

during the diagnostics stage, $ARIMA(1,0,1) \times (1,0,1)_{12}$ is the process that I declare the best specification for forecasting the spread.

4.3.3 *Twelve Month Out-of-Sample Forecast*

I want to conclude the exercise with a forward looking result. I forecast the spread 12 months into the future (until January 2012). Although the preceding forecasts where three-step-ahead predictions I leave the reader with a result that shows the models behaviour over one full seasonal span into the future.

Figure 8
Observed Spread and Predicted Spread by $ARIMA(1,0,1) \times (1,0,1)_{12}$



As prescribed by the forecast evaluation statistics, the process is able to replicate the spread series fairly precisely for the entire period 1997m1 to 2011m1 (see Figure 8). The dashed line, the predicted series, mimics the seasonal pattern of the spread series.

5 *Is unemployment in Canada worse than it looks?*

It is evident from Figure 8 that the spread has jumped to a higher level in the period following the recession that began in the third quarter of 2008. $ARIMA(1,0,1) \times (1,0,1)_{12}$ predicts that the spread will remain at this level for the duration of 2011, following the same seasonal trend exhibited throughout the series. Because ARIMA processes are backward looking, I cannot infer that following this past recession the spread will sustainably remain at this level, mainly because the model is univariate and cannot handle any forward looking influxes of data.

Even though the spread has increased, looking back to when the official UR was at similar levels (see Figure 1); the spread does not seem to be abnormally high. A useful exercise would be to repeat the modelling algorithm presented in this paper for the ratio between the 'real' and official UR's. The resulting series is appropriate to use when one wants to assess whether the number of persons that are marginally attached to the labour force is abnormally high or low.

6 *Conclusion*

In this paper I looked at the spread between the 'real' and official UR's. The spread accounts for individuals that are marginally attached to the labour force (i.e. individuals discouraged by market conditions), and is therefore capable of capturing labour market impacts that follow contractionary periods in the business cycle. To model the future dynamic in the labour market I have developed a modelling algorithm that specifies a top performing model for forecasting the spread. The resulting model was able to impressively replicate the entire spread series and generate accurate predicted values over three month horizons.

I have delivered a model-mining exercise that optimizes the predictive power of a multiplicative seasonal ARIMA process by minimizing out-of-sample forecast errors. For the time series examined, the spread between the real and official unemployment rates, the procedure advocates that out of 576 estimated models, $ARIMA(1,0,1) \times (1,0,1)_{12}$ is not only the best fitted process, but also generates the top performing forecasts. In the case when goodness-of-fit and the forecast performance favour different processes, I advocate using the process that minimizes forecast errors, providing the process generates consistent estimates over time and no serial correlation is present in the residuals. $ARIMA(1,0,1) \times (1,0,1)_{12}$, is ranked first by the Akaike, Bayesian and Hannan-Quinn Information Criterions and produces short term forecasts with an average mean absolute percentage error (MAPE) of 1.76. This supports the consensus that ARIMA processes are capable of producing reliable short-term forecasts. Long-term forecasts should be conducted with caution because ARIMA processes are backward looking in nature (Meyler et al. 1998). A key advantage of the exercise conducted in this paper is that it utilizes objective measures (Information Criterions) to discriminate among processes which results in a semi-automatic algorithm that relies on the user for diagnostics analysis. The modelling algorithm presented is a valid tool for forecasting and optimizing forecast performance over the short-term, and can be applied to any other variable of interest by the analyst.

References

Akyeampong, B. Ernest. 1992. "Discouraged workers – where have they gone?" *Perspectives on Labour and Income.* 4(3). Statistics Canada Catalogue no. 75-001E.

Box, George E.P. and Gwilym M. Jenkins. 1976. "Time Series Analysis: Forecasting and Control." Revised Ed., San Francisco, Holden Day.

Box, George, Gwilym M. Jenkins and Gregory C. Reinsel. 1994. "Time Series Analysis: Forecasting and Control." Prentice-Hall, Inc., 3rd Ed.

Cook, Steven. 2006. "Understanding the construction and interpretation of forecast evaluation statistics using computer-based tutorial exercises." Swansea University. Swansea, Wales, United Kingdom.

Cheung, Yin-Wong and Kon S. Lai. 1995. "Lag Order and Critical Values of the Augmented Dickey-Fuller Test." *Journal of Business & Economic Statistics.* 13(3): 277-280.

Dobre, Ion and Adriana Ana-Maria Alexandru. "Modelling unemployment rate using Box-Jenkins procedure." *Journal of Applied Quantitative Methods.* 3(2): 156-166.

Enders, Walter. 2010. "Applied Econometric Time Series." Hoboken, NJ: John Wiley & Sons, Inc., 3rd Ed.

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

Hamilton, James D. 1994. "Time Series Analysis." Princeton University Press; Princeton, NJ, 1st Ed.

Johnston, Jack and John DiNardo. "Econometric Methods." McGraw-Hill Companies, Inc., 1984, 4th Edition.

Kurita, Takamitsu. 2010. "A Forecasting model for Japan's Unemployment Rate." *Eurasian Journal of Business and Economics.* 3(5): 127-134.

Leitch, Gordon and J. Ernest Tanner. 1991. "Forecast Evaluation: Profits Versus the Conventional Error Measures." *The American Economic Review.* 81(3): 580-590.

Meyler, Aidan, Kenny, Geoff and Terry Quinn. 1998. "Forecasting Irish inflation using ARIMA models." Economic Analysis; Research and Publications Department, Central Bank of Ireland. MPRA Paper No. 11359.

Montgomery, Alan L., Victor Zarnowitz, Ruey S. Tsay and George C. Tiao. 1998. "Forecasting the U.S. Unemployment Rate." *Journal of the American Statistical Association.* 93(442): 478-493.

Poskitt, Don S. and Andrew R. Tremayne. 1987. "Determining a Portfolio of Linear Time Series Models." *Biometrika.* 74(1): 125-137.

Riddell, C. Craig. 1999. "Measuring Unemployment and Structural Unemployment." Department of Economics, University of British Columbia, and Canadian Institute for Advanced Research. Discussion Paper No.: 99-21.

Statistics Canada. 2008. "Methodology of the Canadian Labour Force Survey." Catalogue no. 71-526-X.

Wilson, Patrick J. and Len J. Perry. 2004. "Forecasting Australian Unemployment Rates using Spectral Analysis." *Australian Journal of Labour Economics.* 7(4): 459-480.

Appendix

I.

Ljung – Box Statistics

Table A1

$ARIMA(1,0,1) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	6.0736	0.2991
10	16.4061	0.0886
20	26.3315	0.1551
30	33.1746	0.3150
40	39.7196	0.4828

Table A2

$ARIMA(2,0,0) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	8.5174	0.1299
10	18.8431	**0.0423**
20	31.2429	0.0521
30	38.9800	0.1262
40	44.7685	0.2785

Table A3

$ARIMA(1,0,2) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	7.7474	0.1707
10	26.2929	**0.0034**
20	32.1031	**0.0422**
30	35.7914	0.2150
40	38.1956	0.5517

Table A4

$ARIMA(3,0,0) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	7.0761	0.2150
10	24.6705	**0.0060**
20	29.3684	0.0808
30	32.8239	0.3302
40	35.0390	0.6929

Table A5

$ARIMA(2,0,1) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	10.8454	0.0545
10	20.6732	**0.0235**
20	33.7942	**0.0276**
30	40.4634	0.0962
40	49.3585	0.1474

Ljung – Box Statistics For the top 5 ranked models (by BIC) which have no serially correlated residuals for k=5, 10,20,30,40

Table A6

$ARIMA(2,0,3) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	6.2879	0.2793
10	15.7216	0.1079
20	24.3996	0.2254
30	30.0269	0.4643
40	36.3489	0.6354

Table A7

$ARIMA(4,0,1) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	8.0547	0.1532
10	16.5898	0.0839
20	26.3010	0.1561

30	32.3324	0.3522
40	38.1679	0.5530

Table A8

$ARIMA(1,0,5) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	4.9181	0.4260
10	12.6773	0.2423
20	23.3825	0.2704
30	28.5773	0.5399
40	35.3479	0.6795

Table A9

$ARIMA(3,0,2) \times (1,0,1)_{12}$		
k	$Q(k)$	$Prob > \chi^2(k)$
5	7.3344	0.1969
10	16.0352	0.0986
20	29.8134	0.0729
30	37.5213	0.1624
40	43.5353	0.3234

II.

Model 1 $ARIMA(1,0,1) \times (1,0,1)_{12}$

$$x_t = \frac{\alpha}{(1-\phi_1)(1-\Phi_{12})} + \frac{(1+\theta_1\mathrm{L})(1+\Theta_{12}\mathrm{L}^{12})}{(1-\phi_1\mathrm{L})(1-\Phi_{12}\mathrm{L}^{12})}\varepsilon_t$$

Table A9

	Estimated Coefficient (OPG Std. Error)
$\hat{\phi}_1$	0.9877361* (0.0154184)
$\hat{\Phi}_{12}$	0.9974887* (0.0029907)
$\hat{\theta}_1$	-0.3848386* (0.0778183)
$\hat{\Theta}_{12}$	-0.8103472* (0.1019718)

*Significant at the 0.001 level

Resulting estimated model

$$x_t(1 - 0.9877361\mathrm{L} - 0.9974887\mathrm{L}^{12} + 0.9852556\mathrm{L}^{13})$$

$$= 3.990104 + \varepsilon_t(1 - 0.3848386\mathrm{L} - 0.8103472\mathrm{L}^{12} + 0.3113853\mathrm{L}^{13})$$

III.

Model 1 $ARIMA(1,0,1) \times (1,0,1)_{12}$

Solving for the roots of polynomial $A(L)$

$$\lambda^{13} - 0.9877361\lambda^{12} - 0.9974887\lambda + 0.9852556 = 0$$

yields

Table A10

λ
± 0.99979
0.987737
$-0.865844 \pm 0.499895i$
$-0.499895 \pm 0.865844i$
$-4.28772 \times 10^{-9} \pm 0.99979i$
$0.499895 \pm 0.865844i$
$0.865844 \pm 0.499895i$

All λ's lie outside the unit circle (i.e. $\forall|\lambda| < 1$),

IV.

Table A11
Forecast Evaluation Statistics

$\boldsymbol{ME} = \frac{1}{T}\sum_{t=1}^{T} e_t$	$\boldsymbol{MAE} = \frac{1}{T}\sum_{t=1}^{T}\lvert e_t\rvert$	$\boldsymbol{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2}$
$\boldsymbol{MPE} = \frac{100}{T}\sum_{t=1}^{T}\frac{e_t}{y_t}$	$\boldsymbol{MAPE} = \frac{100}{T}\sum_{t=1}^{T}\frac{\lvert e_t\rvert}{y_t}$	$\boldsymbol{TU} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} y_t^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T} f_t^2}}$